01/28/02 J1064 U.S. PTO

01-31-02 PCT

JC10 Rec'd PCT/PTO 26 JAN 2002

FORM PTO-1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER
PF-0728 USN

U.S. APPLICATION NO. (If known, see 37 CFR 1.5)
TO BE ASSIGNED 10/049742

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/US00/21313	**03 August 2000**	**03 August1999**

TITLE OF INVENTION
HUMAN CHAPERONE PROTEINS

APPLICANT(S) FOR DO/EO/US
YUE, Henry; BANDMAN, Olga; TANG, Y. Tom; BAUGHN, Mariah R.; AZIMZAI, Yalda; LU, Dyung Aina M.

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is the **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☐ This is an express request to promptly begin national examination procedures (35 U.S.C. 371 (f)).
4. ☐ The US has been elected by the expiration of 19 months from the priority date (PCT Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☐ is attached hereto (required only if not communicated by the International Bureau)
   b. ☐ has been communicated by the International Bureau.
   c. ☒ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☐ have not been made and will not be made.
   e. ☐ attached hereto Article 34 Amendment
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 16 below concern document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☒ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.27 and 3.31 is included.
13. ☒ A FIRST preliminary amendment, as follows: Cancel in this application original claims 12, 14, 18, 20, 21, 23, 24 & 27 before calculating the filing fee, without prejudice or disclaimer. Applicants submit that these claims were included in the application as filed in the interest of providing notice to the public of certain specific subject matter intended to be claimed, and are being canceled at this time in the interest of reducing filing costs. Applicants expressly state that these claims are not being canceled for reasons related to patentability, and are in fact fully supported by the specification as filed. Applicants expressly reserve the right to reinstate these claims or to add other claims during prosecution of this application or a continuation or divisional application. Applicants expressly do not disclaim the subject matter of any invention disclosed herein which is not set forth in the instantly filed claims.
    ☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter.
16. ☒ Other items or information:

1) Transmittal Letter (2 pp, in duplicate)
2) Return Postcard
3) Express Mail Label No.: EL 856 146 666 US
4) Sequence Listing Statement

JC13 Rec'd PCT/PTO 26 JAN 2002

U.S. APPLICATION NO. (if known, see 37 CFR 1.5) TO BE ASSIGNED 10/049742	INTERNATIONAL APPLICATION NO.: PCT/US00/21313	ATTORNEY'S DOCKET NUMBER **PF-0728 USN**

17. ☒ The following fees are submitted:

BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5):

Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO.......$1000.00

☐ International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO..$860.00

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO........$710.00

☒ **International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4)..........$710.00**

☐ International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4)..................$100.00

ENTER APPROPRIATE BASIC FEE AMOUNT =				$710.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).				$	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total Claims	20 =	0	**X $ 18.00**	$	
Independent Claims	2 =	0	**X $ 80.00**	$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			**+ $270.00**	$	
TOTAL OF ABOVE CALCULATIONS =				$	
☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.				$	
SUBTOTAL =				$710.00	
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1492(f)). +				$	
TOTAL NATIONAL FEE =				$710.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by the appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +					
TOTAL FEES ENCLOSED =				$710.00	
				Amount to be Refunded:	$
				Charged:	$

a. ☐ A check in the amount of $________ to cover the above fees is enclosed.
b. ☒ Please charge my Deposit Account No. 09-0108 in the amount of $710.00 to cover the above fees.
c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 09-0108. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

INCYTE GENOMICS, INC.
3160 Porter Drive
Palo Alto, CA 94304

SIGNATURE

NAME: Diana Hamlet-Cox

REGISTRATION NUMBER: 33,302

DATE: 28 January 2002

HUMAN CHAPERONE PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human chaperone proteins and to the use of these sequences in the diagnosis, treatment, and prevention of reproductive, eye, neuromuscular, metabolic, and autoimmune/inflammatory disorders, infectious diseases, and cell proliferative disorders including cancer.

BACKGROUND OF THE INVENTION

Molecular chaperones are proteins that interact with many other cellular proteins. Chaperones are involved in normal cellular functions, such as the folding of newly synthesized polypeptides and the assembly of multisubunit protein structures, in the transport of proteins across membranes, and in the stabilization of proteins in inactive configurations. This may give chaperones a role in cell signaling, as they can limit the access of proteins to their signaling partners, and in any regulatory process dependent on oligomerization or complex protein rearrangements. Chaperones are also involved in cellular responses to stresses such as toxicity and heat shock, and are therefore called heat-shock proteins (Hsp).

Chaperones are found in many cellular compartments. In the mitochondria, for example, chaperones on both sides of the membrane are involved in importing most of the nuclear encoded proteins necessary for oxidative phosphorylation. Chaperones may be divided into several classes named for their approximate molecular weights, including Hsp90, Hsp70,Hsp 60, Hsp40 (also called DnaJ), and the small Hsps (having molecular masses between 20 kD and 30 kD). Mitochondrial chaperones show a high degree of similarity to molecular chaperones in bacteria, and in general chaperones are ubiquitous and highly conserved, from bacteria to humans (Martinus, R.D. et al. (1995) FASEB J. 9(5):371-378).

Molecular chaperone genes are activated by a variety of stresses, including glucose deprivation, ethanol, and heavy metals as well as heat shock, all of which affect protein folding and aggregation. Activation may be expected in any disorder that results in temperature elevation. Molecular chaperones have been suggested to play a role in the development of autoimmune conditions and have been implicated in a variety of metabolic and developmental disorders as well as in response to trauma. In addition, because many of the proteins that carry out the major mitochondrial function, oxidative phosphorylation, must be imported from the cytoplasm, any disorder affecting metabolism may involve mitochondrial import translocases.

Under normal or nonstressed conditions, constitutively expressed Hsps facilitate proper protein

folding and maturation, promote protein translocation across membranes, and regulate hormone receptor and protein kinase activity (Hightower, L.E. et al. (1991) Cell 66:191-197), antigen presentation, protein degradation in the lysosome, and uncoating of clathrin-coated vesicles. Hsps are located in all major cellular compartments and function as monomers, multimers, or in complexes with other cellular proteins, which may determine the rate and specificity of Hsp action. Different roles have been ascribed to different classes of Hsps. Hsp20 proteins seem to form heterooligomers that can protect other proteins against heat-induced denaturation and aggregation. Hsp40, homologous to the bacterial DnaJ protein, and Hsp70, homologous to bacterial DnaK, act in concert to aid in protein folding and assembly of higher order protein complexes. Hsp60, along with Hsp10, binds misfolded proteins and gives them the opportunity to refold correctly (Alberts, B. et al. (1994) Molecular Biology of the Cell Garland Publishing Co., New York, NY p. 608).

Hsp70 is a dimeric and ubiquitous protein which binds its substrates in an extended conformation through hydrophobic interactions. Hsp70 binds to newly synthesized proteins and is required for protein transport. The strength and specificity of Hsp70's interaction with its substrates is modified by binding and hydrolysis of ATP. Hsp70 has low protein affinity in its ATP-bound state, and increased protein affinity after ATP is hydrolyzed to ADP (Burston, S.G. and Clarke, A.R. (1995) Essays Biochem. 29:125-136). DnaJ chaperones work in concert with Hsp70. In particular, DnaJ interacts with the ATPase domain of Hsp70. The defining characteristic of the DnaJ chaperone family is an N-terminal, approximately 70 amino acid signature called the J domain, which is required for interactions with Hsp70. The tripeptide HPD seems to be particularly important for this interaction (Kelley, W. (1999) Curr. Biol. (1999) 9:R305-308). DnaJ stimulates ATP hydrolysis, increasing the affinity of Hsp70 for its protein substrate. GrpE, another co-chaperone, promotes dissociation of ADP from Hsp70, again modifying the Hsp70/substrate interaction and completing the cycle (Burston and Clarke, supra). Many eukaryotic DnaJ homologs have recently been described. Growing evidence suggests that specific DnaJ homologs interact with specific Hsp70 homologs to form a chaperone complex with affinity for specific substrates (Kelley, W. (1998) Trends in Biochem. Sci. 23:222-227). The DnaJ homolog Hsp40 was shown to co-localize with Hsp70 in the nuclei and nucleoli of heat-shocked HeLa cells (Ohtsuka, K. (1993) Biochem. Biophys. Res. Commun. 197:235-240). Homologs of GrpE have been identified in bovine, porcine, and rat tissues (Naylor, D.J. et al. (1995) Biochim. Biophys. Acta 1248:75-79).

The induction of heat shock proteins (Hsps), is a physiological and biochemical response to abrupt increases in temperature or exposure to a variety of other metabolic insults including heavy metals, amino acid analogs, toxins, and oxidative stress. This response occurs in all prokaryotic and eukaryotic cells and is characterized by repression of normal protein synthesis and initiation of

transcription of Hsp-encoding genes (Hightower, supra). During cellular stress, Hsps form a complex with proteins that misfold or unfold, either "rescuing" these proteins from irreversible damage or increasing their susceptibility to proteolytic attack. Overexpression of Hsps in transgenic mice and rats, or heat treatment of normal animals to induce Hsps, protects the heart muscle from ischemic injury. Both heat shock-induced and exogenous Hsps protect smooth muscle cells from serum deprivation-induced cell death. Overexpression of Hsps also protects murine fibroblasts from both UV light injury and proinflammatory cytokines released during UV exposure (Marber, M.S. et al. (1995) J. Clin. Invest. 95:1446-1456; Simon, M.M. et al. (1995) J. Clin. Invest. 95:926-933). Specific Hsps bind immunosuppressive drugs and may play a role in modulation of immune responses. Hsps expressed in cancer cells can protect the cancer cells from the cytotoxic effects of drugs used in anticancer therapies. Purified Hsps isolated from tumor cells and used as antigens have been shown to provide immunity to the tumors from which they are isolated (Udono, H. et al. (1994) J. Immunol. 152:5398-5403; Young R.A. (1990) Annu. Rev. Immunol. 8:401-420; Marber, M.S. et al. (1995) J. Clin. Invest. 95:1446-1456; Simon, M.M. et al. (1995) J. Clin. Invest. 95:926-933).

Chaperones are useful as markers of environmental stress and disease, and are associated with a variety of diseases and immune and drug responses. Several of the constitutive Hsp genes are located in the major histocompatibility complex on chromosome 6. Members of the Hsp family have also been shown to play roles in T-cell mediated regulation of inflammation and immune recognition. For example, Hsp90 binds to steroid hormone receptors, represses transcription in the absence of the ligand, and provides the proper folding of the ligand-binding domain in the presence of the hormone (Burston and Clarke, supra). Heat shock treatment of B-cells enhances processing of antigen and the assembly and function of MHC class II molecules (Sargent, C.A. et al. (1989) Proc. Natl. Acad. Sci. USA 86:1968-1972; Fang, Y. et al. (1996) J. Biol. Chem. 271:28697-28702; Hendrick, J.P. et al. (1993) Proc. Natl. Acad. Sci. USA 90:10216-10220). Abnormal transcription of Hsp70 has been associated with major depression (Shimizu, S. et al. (1996) Biochem. Biophys. Res. Commun. 219:745-752). Hsp70 expression increases in response to tobacco smoke (Vayssier, M. et al. (1998) Biochem. Biophys. Res. Commun. 252:249-256). Hsp70 is involved in drug resistance in breast cancer patients treated with combination chemotherapies (Vargas-Roig, L.M. et al. (1998) Int. J. Cancer 79:468-475). Hsp70 variants are associated with clozapine-induced agranulocytosis, an adverse drug reaction (Turbay, D. et al. (1997) Blood 89:4167-4174). Knockout mice have provided additional information on the roles of Hsp70 in reproduction. For example, female homozygous knockout mice for Hsp70 are found to undergo normal meiosis and are fertile. In contrast, the homozygous male knockout mice lack postmeiotic spermatids and mature sperm, and are infertile (Dix, D.J. et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:3264-3268). A DnaJ (Hsp40) homolog is essential for normal placental development in

mice (Hunter, P. et al. (1999) Development 126:1247-1258). Other DnaJ homologs have been implicated in viral DNA replication, secretion, tumor suppression, microtubule formation in M phase, and influenza virus infection (Kelley, 1998, supra).

The small Hsps are able to supress aggregation and heat inactivation of various proteins, including actin (Hickey, E. et al. (1986) Nucleic Acids Res. 14:4127-4145; Miron, T. et al. (1991) J. Cell Biol. 114:255-261). α-Crystallin, a protein abundant in the lens of the eye, is an oligomer of two subunits, αA and αB, which are 55% identical and belong to the small Hsp family. α-Crystallin is thought to be important for maintaining the transparency of the lens by preventing denaturation and aggregation of proteins. A missense mutation in the αA-crystallin gene is associated with autosomal dominant congenital cataracts (Litt, M. et al. (1998) Hum. Molec. Genet. 7:471-474). However, the functional role of α-crystallin is not confined to the eye. A missense mutation of αB-crystallin has been shown to cause a desmin-related myopathy (Vicart, P. et al. (1998) Nat. Genet. 20:92-95). Desmin-related myopathies are inherited neuromuscular disorders.

The discovery of new human chaperone proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of reproductive, eye, neuromuscular, metabolic, and autoimmune/inflammatory disorders, infectious diseases, and cell proliferative disorders including cancer.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, human chaperone proteins, referred to collectively as "HCPN" and individually as "HCPN-1," "HCPN-2," "HCPN-3," "HCPN-4," "HCPN-5," "HCPN-6," "HCPN-7," "HCPN-8," "HCPN-9," "HCPN-10," and "HCPN-11." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-11.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of

SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-11. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:12-22.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence having at least 70% sequence

identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and a pharmaceutically acceptable excipient. In one

embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional HCPN, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional HCPN, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional HCPN, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid

sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:12-22, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID

NO:12-22, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of the above polynucleotide sequence; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding HCPN.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of HCPN.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding HCPN were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although

any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"HCPN" refers to the amino acid sequences of substantially purified HCPN obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of HCPN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of HCPN either by directly interacting with HCPN or by acting on components of the biological pathway in which HCPN participates.

An "allelic variant" is an alternative form of the gene encoding HCPN. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding HCPN include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as HCPN or a polypeptide with at least one functional characteristic of HCPN. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding HCPN, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HCPN. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent HCPN. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of HCPN is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may

include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of HCPN. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of HCPN either by directly interacting with HCPN or by acting on components of the biological pathway in which HCPN participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind HCPN polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having

modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic HCPN, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding HCPN or fragments of HCPN may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of HCPN or the polynucleotide encoding HCPN which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid

residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:12-22 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:12-22, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:12-22 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:12-22 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:12-22 and the region of SEQ ID NO:12-22 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-11 is encoded by a fragment of SEQ ID NO:12-22. A fragment of SEQ ID NO:1-11 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-11. For example, a fragment of SEQ ID NO:1-11 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-11. The precise length of a fragment of SEQ ID NO:1-11 and the region of SEQ ID NO:1-11 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191.

For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in

a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of HCPN which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of HCPN which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of HCPN. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of HCPN.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs

preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an HCPN may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of HCPN.

"Probe" refers to nucleic acid sequences encoding HCPN, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South

West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and

other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding HCPN, or fragments thereof, or HCPN itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells

includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook, J. et al. (1989), supra.

A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999)

set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human chaperone proteins (HCPN), the polynucleotides encoding HCPN, and the use of these compositions for the diagnosis, treatment, or prevention of reproductive, eye, neuromuscular, metabolic, and autoimmune/inflammatory disorders, infectious diseases, and cell proliferative disorders including cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding HCPN. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each HCPN were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. In some cases, GenBank sequence identifiers are also shown in column 5. The Incyte clones and GenBank cDNA sequences, where indicated, in column 5 were used to assemble the consensus nucleotide sequence of each HCPN and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding HCPN. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:12-22 and to distinguish between SEQ ID NO:12-22 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express HCPN as a fraction of total tissues expressing HCPN. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing HCPN as a fraction of total tissues expressing HCPN. Column 5 lists the vectors used to subclone each cDNA library.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries

from which cDNA clones encoding HCPN were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

SEQ ID NO:14 maps to chromosome 10 within the interval from 46.2 to 46.8 centiMorgans. This interval also contains a gene with homology to the murine leukemia viral (bmi-1) oncogene. SEQ ID NO:15 maps to chromosome 16 within the interval from 50.8 to 56.2 centiMorgans. This interval also contains a gene associated with neuronal ceroid lipofuscinosis, also known as Batten disease. SEQ ID NO:22 maps to chromosome 1 within the interval from 78.3 to 84.2 centiMorgans, to chromosome 6 within the interval from 91.8 to 96.1 centiMorgans, to chromosome 10 within the interval from 93.8 to 96.9 centiMorgans, and to chromosome 12 within the interval from 13.8 to 24.6 centiMorgans. The interval on chromosome 1 from 78.3 to 84.2 centiMorgans also contains genes and/or ESTs associated with myopathy, hypoketotic hypoglycemia and hyperthyroxinemia. The interval on chromosome 6 from 91.8 to 96.1 centiMorgans also contains a gene and/or EST associated with maple syrup urine disease. The interval on chromosome 12 from 13.8 to 24.6 centiMorgans also contains genes and/or ESTs associated with T cell antigen T4 deficiency, neonatal adrenoleukodystrophy, and Zellweger syndrome.

The invention also encompasses HCPN variants. A preferred HCPN variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the HCPN amino acid sequence, and which contains at least one functional or structural characteristic of HCPN.

The invention also encompasses polynucleotides which encode HCPN. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:12-22, which encodes HCPN. The polynucleotide sequences of SEQ ID NO:12-22, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding HCPN. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding HCPN. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:12-22 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:12-22. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of HCPN.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding HCPN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring HCPN, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode HCPN and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring HCPN under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HCPN or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HCPN and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode HCPN and HCPN derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HCPN or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:12-22 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (PE Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is

automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (PE Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (PE Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding HCPN may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode HCPN may be cloned in recombinant DNA molecules that direct expression of HCPN, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express HCPN.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HCPN-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of HCPN, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby

maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding HCPN may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, HCPN itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of HCPN, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active HCPN, the nucleotide sequences encoding HCPN or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding HCPN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HCPN. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding HCPN and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression

vectors containing sequences encoding HCPN and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding HCPN. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al., (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding HCPN. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding HCPN can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding HCPN into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem.

264:5503-5509.) When large quantities of HCPN are needed, e.g. for the production of antibodies, vectors which direct high level expression of HCPN may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of HCPN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, supra; and Scorer, supra.)

Plant systems may also be used for expression of HCPN. Transcription of sequences encoding HCPN may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, supra; Broglie, supra; and Winter, supra.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding HCPN may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses HCPN in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of HCPN in cell lines is preferred. For example, sequences encoding HCPN can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before

being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk⁻* and *apr⁻* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding HCPN is inserted within a marker gene sequence, transformed cells containing sequences encoding HCPN can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HCPN under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding HCPN and that express HCPN may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of HCPN using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence

activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HCPN is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HCPN include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HCPN, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding HCPN may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HCPN may be designed to contain signal sequences which direct secretion of HCPN through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid

sequences encoding HCPN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric HCPN protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of HCPN activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the HCPN encoding sequence and the heterologous protein sequence, so that HCPN may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled HCPN may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

HCPN of the present invention or fragments thereof may be used to screen for compounds that specifically bind to HCPN. At least one and up to a plurality of test compounds may be screened for specific binding to HCPN. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of HCPN, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which HCPN binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express HCPN, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing HCPN or cell membrane fractions which contain HCPN are then contacted with a test compound and binding, stimulation, or inhibition of activity of either HCPN or the

compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with HCPN, either in solution or affixed to a solid support, and detecting the binding of HCPN to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

HCPN of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of HCPN. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for HCPN activity, wherein HCPN is combined with at least one test compound, and the activity of HCPN in the presence of a test compound is compared with the activity of HCPN in the absence of the test compound. A change in the activity of HCPN in the presence of the test compound is indicative of a compound that modulates the activity of HCPN. Alternatively, a test compound is combined with an in vitro or cell-free system comprising HCPN under conditions suitable for HCPN activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of HCPN may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding HCPN or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential

therapeutic or toxic agents.

Polynucleotides encoding HCPN may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding HCPN can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding HCPN is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress HCPN, e.g., by secreting HCPN in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of HCPN and human chaperone proteins. In addition, the expression of HCPN is closely associated with reproductive disorders and with cancerous, proliferating, inflamed, and hematopoietic/immune tissues. Therefore, HCPN appears to play a role in reproductive, eye, neuromuscular, metabolic, and autoimmune/inflammatory disorders, infectious diseases, and cell proliferative disorders including cancer. In the treatment of disorders associated with increased HCPN expression or activity, it is desirable to decrease the expression or activity of HCPN. In the treatment of disorders associated with decreased HCPN expression or activity, it is desirable to increase the expression or activity of HCPN.

Therefore, in one embodiment, HCPN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HCPN. Examples of such disorders include, but are not limited to, a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a disorder of the eye such as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma,

amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retrobulbar tumor, and chiasmal tumor; a neuromuscular disorder such as a desmin-related myopathy; a metabolic disorder such as Zellweger syndrome, maple syrup urine disease, adrenoleukodystropy, carnitine palmitoyltransferase deficiency, Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria; an autoimmune/inflammatory disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a viral infection, such as those caused by adenoviruses (acute respiratory disease, pneumonia), arenaviruses (lymphocytic choriomeningitis), bunyaviruses (Hantavirus), coronaviruses (pneumonia, chronic bronchitis), hepadnaviruses (hepatitis), herpesviruses (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), flaviviruses (yellow fever), orthomyxoviruses (influenza), papillomaviruses (cancer), paramyxoviruses (measles, mumps), picornoviruses (rhinovirus, poliovirus, coxsackie-virus), polyomaviruses (BK virus, JC virus), poxviruses (smallpox), reovirus (Colorado tick fever),

retroviruses (human immunodeficiency virus, human T lymphotropic virus), rhabdoviruses (rabies), rotaviruses (gastroenteritis), and togaviruses (encephalitis, rubella); and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing HCPN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HCPN including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified HCPN in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HCPN including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of HCPN may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HCPN including, but not limited to, those listed above.

In a further embodiment, an antagonist of HCPN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of HCPN. Examples of such disorders include, but are not limited to, those reproductive, eye, neuromuscular, metabolic, and autoimmune/inflammatory disorders, infectious diseases, and cell proliferative disorders, including cancer, described above. In one aspect, an antibody which specifically binds HCPN may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express HCPN.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HCPN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of HCPN including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with

lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of HCPN may be produced using methods which are generally known in the art. In particular, purified HCPN may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HCPN. Antibodies to HCPN may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with HCPN or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HCPN have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of HCPN amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to HCPN may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HCPN-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be

generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for HCPN may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between HCPN and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering HCPN epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for HCPN. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of HCPN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple HCPN epitopes, represents the average affinity, or avidity, of the antibodies for HCPN. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular HCPN epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the HCPN-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of HCPN, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine

the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of HCPN-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding HCPN, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding HCPN. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding HCPN. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding HCPN may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii)

express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in HCPN expression or regulation causes disease, the expression of HCPN from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in HCPN are treated by constructing mammalian expression vectors encoding HCPN and introducing these vectors by mechanical means into HCPN-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of HCPN include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). HCPN may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding HCPN from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method

(Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to HCPN expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding HCPN under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding HCPN to cells which have one or more genetic abnormalities with respect to the expression of HCPN. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544; and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver

polynucleotides encoding HCPN to target cells which have one or more genetic abnormalities with respect to the expression of HCPN. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing HCPN to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding HCPN to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotech. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for HCPN into the alphavirus genome in place of the capsid-coding region results in the production of a large number of HCPN-coding RNAs and the synthesis of high levels of HCPN in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of HCPN into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones

of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HCPN.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HCPN. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages

within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding HCPN. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased HCPN expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding HCPN may be therapeutically useful, and in the treament of disorders associated with decreased HCPN expression or activity, a compound which specifically promotes expression of the polynucleotide encoding HCPN may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding HCPN is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding HCPN are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding HCPN. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids

Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of HCPN, antibodies to HCPN, and mimetics, agonists, antagonists, or inhibitors of HCPN.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of

an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising HCPN or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, HCPN or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HCPN or fragments thereof, antibodies of HCPN, and agonists, antagonists or inhibitors of HCPN, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and

methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind HCPN may be used for the diagnosis of disorders characterized by expression of HCPN, or in assays to monitor patients being treated with HCPN or agonists, antagonists, or inhibitors of HCPN. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for HCPN include methods which utilize the antibody and a label to detect HCPN in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring HCPN, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of HCPN expression. Normal or standard values for HCPN expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to HCPN under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of HCPN expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding HCPN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of HCPN may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of HCPN, and to monitor regulation of HCPN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HCPN or closely related molecules may be used to identify nucleic acid sequences which encode HCPN. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding HCPN, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the HCPN encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:12-22 or from genomic sequences including promoters, enhancers, and introns of the HCPN gene.

Means for producing specific hybridization probes for DNAs encoding HCPN include the cloning of polynucleotide sequences encoding HCPN or HCPN derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding HCPN may be used for the diagnosis of disorders associated with expression of HCPN. Examples of such disorders include, but are not limited to, a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a disorder of the eye such as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma, amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retrobulbar tumor, and chiasmal tumor; a neuromuscular disorder such as a desmin-related myopathy; a metabolic disorder such as Zellweger syndrome, maple syrup urine disease, adrenoleukodystropy, carnitine palmitoyltransferase deficiency, Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria; an autoimmune/inflammatory disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,

allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a viral infection, such as those caused by adenoviruses (acute respiratory disease, pneumonia), arenaviruses (lymphocytic choriomeningitis), bunyaviruses (Hantavirus), coronaviruses (pneumonia, chronic bronchitis), hepadnaviruses (hepatitis), herpesviruses (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), flaviviruses (yellow fever), orthomyxoviruses (influenza), papillomaviruses (cancer), paramyxoviruses (measles, mumps), picornoviruses (rhinovirus, poliovirus, coxsackie-virus), polyomaviruses (BK virus, JC virus), poxviruses (smallpox), reovirus (Colorado tick fever), retroviruses (human immunodeficiency virus, human T lymphotropic virus), rhabdoviruses (rabies), rotaviruses (gastroenteritis), and togaviruses (encephalitis, rubella); and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding HCPN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered HCPN expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding HCPN may be useful in assays that

detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding HCPN may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding HCPN in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of HCPN, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding HCPN, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding HCPN may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding HCPN, or a fragment of a polynucleotide complementary to the polynucleotide encoding HCPN, and will be employed under optimized conditions for identification of a specific gene or condition.

Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding HCPN may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding HCPN are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of HCPN include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor

progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, antibodies specific for HCPN, or HCPN or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression <u>in vivo</u>, as in the case of a tissue or biopsy sample, or <u>in vitro</u>, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with <u>in vitro</u> model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the

rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for HCPN to quantify the levels of HCPN expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in <u>DNA Microarrays: A Practical Approach</u>, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding HCPN may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding HCPN on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, HCPN, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes

between HCPN and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with HCPN, or fragments thereof, and washed. Bound HCPN is then detected by methods well known in the art. Purified HCPN can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding HCPN specifically compete with a test compound for binding HCPN. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HCPN.

In additional embodiments, the nucleotide sequences which encode HCPN may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/146,908 and U.S. Ser. No. 60/160,924, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (PE Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened

for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:12-22. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{\text{5 x minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and

70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding HCPN occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Chromosomal Mapping of HESHP Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:12-22 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:12-22 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

The genetic map locations of SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:22 are described in The Invention as ranges, or intervals, of human chromosomes. More than one map location is reported for SEQ ID NO:22, indicating that previously mapped sequences having similarity, but not complete identity, to SEQ ID NO:22 were assembled into their respective clusters. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Diseases associated with the public and Incyte sequences located within the indicated intervals are also reported in the Invention where applicable.

VI. Extension of HCPN Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:12-22 were produced by extension of an

appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing

media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

In like manner, the polynucleotide sequences of SEQ ID NO:12-22 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such extension, and an appropriate genomic library.

VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:12-22 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing

photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is

then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60 °C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65 °C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60 °C. The arrays are washed for 10 min at 45 °C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45 °C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

IX. Complementary Polynucleotides

Sequences complementary to the HCPN-encoding sequences, or any parts thereof, are used to

detect, decrease, or inhibit expression of naturally occurring HCPN. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of HCPN. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the HCPN-encoding transcript.

X. Expression of HCPN

Expression and purification of HCPN is achieved using bacterial or virus-based expression systems. For expression of HCPN in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express HCPN upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of HCPN in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding HCPN by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, HCPN is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from HCPN at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins

(QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified HCPN obtained by these methods can be used directly in the assays shown in Examples XI and XV.

XI. Demonstration of HCPN Activity

HCPN induction by heat or toxins may be demonstrated using primary cultures of human fibroblasts or human cell lines such as CCL-13, HEK293, or HEP G2 (ATCC). To heat induce HCPN expression, aliquots of cells are incubated at 42°C for 15, 30, or 60 minutes. Control aliquots are incubated at 37 °C for the same time periods. To induce HCPN expression by toxins, aliquots of cells are treated with 100 µM arsenite or 20 mM azetidine-2-carboxylic acid for 0, 3, 6, or 12 hours. After exposure to heat, arsenite, or the amino acid analogue, samples of the treated cells are harvested and cell lysates prepared for analysis by western blot.

Cells are lysed in lysis buffer containing 1% Nonidet P-40, 0.15 M NaCl, 50 mM Tris-HCl, 5 mM EDTA, 2 mM N-ethylmaleimide, 2 mM phenylmethylsulfonyl fluoride, 1 mg/ml leupeptin, and 1 mg/ml pepstatin. Twenty micrograms of the cell lysate is separated on an 8% SDS-PAGE gel and transferred to a nitrocellulose membrane. After blocking with 5% nonfat dry milk/phosphate-buffered saline for 1 h, the membrane is incubated overnight at 4°C or at room temperature for 2-4 hours with a 1:1000 dilution of anti-HCPN serum in 2% nonfat dry milk/phosphate-buffered saline. The membrane is then washed and incubated with a 1:1000 dilution of horseradish peroxidase-conjugated goat anti-rabbit IgG in 2% dry milk/phosphate-buffered saline. After washing with 0.1% Tween 20 in phosphate-buffered saline, the HCPN protein is detected and compared to controls by using chemiluminescence. Induction of HCPN under stress conditions is evidence of HCPN activity.

Alternatively, HCPN activity can be determined by measuring the ability to promote ATP hydrolysis by Hsp70. Briefly, 1 µg Hsp70 protein is incubated with 1 nmol unlabled ATP and 0.01 µCi of α32P-ATP in ATPase buffer (50 mM HEPES, pH 7.4, 50 mM NaCl, 10 mM DTT, and 2 mM $MgCl_2$) in a total volume of 20 µl at 30°C with or without HCPN. After 1 hr, 1 µl of the reaction is spotted on polyethyleneimine cellulose TLC plates and developed in 1M formic acid with 0.5 M LiCl. Plates are examined for conversion of α32P-ATP to α32P-ADP by phosphorimager (Hunter, supra).

XII. Functional Assays

HCPN function is assessed by expressing the sequences encoding HCPN at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome

formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of HCPN on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding HCPN and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding HCPN and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIII. Production of HCPN Specific Antibodies

HCPN substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the HCPN amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St.

Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-HCPN activity by, for example, binding the peptide or HCPN to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIV. Purification of Naturally Occurring HCPN Using Specific Antibodies

Naturally occurring or recombinant HCPN is substantially purified by immunoaffinity chromatography using antibodies specific for HCPN. An immunoaffinity column is constructed by covalently coupling anti-HCPN antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HCPN are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HCPN (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HCPN binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HCPN is collected.

XV. Identification of Molecules Which Interact with HCPN

HCPN, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HCPN, washed, and any wells with labeled HCPN complex are assayed. Data obtained using different concentrations of HCPN are used to calculate values for the number, affinity, and association of HCPN with the candidate molecules.

Alternatively, molecules interacting with HCPN are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

HCPN may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be

understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	12	723593	SYNOOAT01	723593H1 (SYNOOAT01), 723593R1 (SYNOOAT01), 1722150F6 (BLADNOT06), 1955474F6 (CONNNOT01), 2480728F6 (SMCANOT01), 3873248H1 (HEARNOT06), 3961279H2 (HEARFET02)
2	13	1708350	PROSNOT16	1708350H1 (PROSNOT16), 1998534T6 (BRSTTUT03), 2613134F6 (BRSTTUT14)
3	14	1742550	HIPONON01	599920F1 (BRSTNOT02), 599920R1 (BRSTNOT02), 960276H1 (BRSTTUT03), 1793990H1 (PROSTUT05), 1946779R6 (PITUNOT01), 1998420R6 (BRSTTUT03), 2364051F6 (ADRENOT07), 3819195H1 (BONSTUT01), g1118481
4	15	1919301	BRSTTUT01	1359330F1 (LUNGNOT12), 1430543F6 (SINTBST01), 1919301H1 (BRSTTUT01), 2120689T6 (BRSTTUT02), 2129362R6 (KIDNNOT05)
5	16	2012055	TESTNOT03	2012055H1 (TESTNOT03), 2012055R6 (TESTNOT03), 2012055T6 (TESTNOT03)
6	17	2238062	PANCTUT02	2238062H1 (PANCTUT02), 3456711H1 (293TF1T01), 3660019H1 (ENDPNOT02), 6026459H1 (TESTNOT11), SAEB00269R1, SAEB00917F1
7	18	1459372	COLNFET02	1459372F1 (COLNFET02), 1459372H1 (COLNFET02), 1930568F6 (COLNTUT03), 2917949F6 (THYMFET03)
8	19	1825012	LSUBNOT03	183307R6 (CARDNOT01), 306328H1 (HEARNOT01), 636856H1 (NEUTGMT01), 954844R1 (KIDNNOT05), 1406982F6 (LATRTUT02), 1675115F6 (BLADNOT05), 1735613F6 (COLNNOT22), 1825012F6 (LSUBNOT03), 1825012H1 (LSUBNOT03), 2642713H1 (LUNGTUT08), 2690551H1 (LUNGNOT23), 2734283T6 (OVARTUT04), 2984675H1 (CARGDIT01), 3234716F6 (COLNUCT03), 3732983F6 (SMCCNOS01), 3732983T6 (SMCCNOS01), 5863484H1 (MUSLTDT01), SAUA01247F1, SAUB00235F1, SAUA01545F1
9	20	1906464	OVARNOT07	1906464H1 (OVARNOT07), 1906464T6 (OVARNOT07), g2054718
10	21	1979146	LUNGTUT03	1232212T1 (LUNGFET03), 1979146H1 (LUNGTUT03), 2524756F6 (BRAITUT21)
11	22	5680480	BRAENOT02	757593H1 (BRAITUT02), 1349105T6 (LATRTUT02), 1645668H1 (PROSTUT09), 1803569T6 (SINTNOT13), 3179743H1 (TLYJNOT01), 3659193H1 (ENDPNOT02), 5016517H1 (BRAXNOT03), 5173162H1 (EPIBTXT01), 5680480H1 (BRAENOT02), g1148507

Table 2

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
1	170	T77 S92 S2 T8 T148 T163		Alpha crystallin signature: A67-I149, V132-R153 Heat shock HSP20 signature: F55-Q162	Heat shock protein [Mus musculus] g6580603	BLAST-GenBank BLAST-DOMO BLIMPS-PRINTS HMMER-PFAM MOTIFS
2	304	S112 S245 S269 T86 S120 T170 S210 S211 T223		DnaJ signature: D11-R76, A22-D41, R32-E91 Nt-DnaJ domains: L12-R85, E26-K42, F53-D73	Protein containing similarity to a DnaJ-like domain [Caenorhabditis elegans] g2384916	BLAST-GenBank BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS ProfileScan
3	483	S18 S79 T123 S6 T194 T196 T218 T232 S233 T260 S292 T334 T341 S408 S409 S421 S460 S113 S136 T163 T251 T297 T383 S413 S446 Y217	N249 N406	DnaJ signature: M1-L58, A5-D24, R15-D64 Nt-DnaJ domains: D9-K25, F35-D55, K3-V93 DnaJ homolog MTJ1: D56-T232 Transmembrane domain: L85-Y103	DnaJ-like protein [Mus musculus] g473847	BLAST-GenBank BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER HMMER-PFAM MOTIFS ProfileScan
4	226	S39 S116 T105 S137 T144 S197	N158	DnaJ signature: A49-L114, A60-D79, A68-P133 Nt-DnaJ domains: Q64-R80, T48-R112, F91-D111 Signal peptide: M1-A28 Transmembrane domain: D208-I226	TCJ4 (DnaJ family of protein chaperones) [Trypanosoma cruzi] g960294	BLAST-GenBank BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER HMMER-PFAM MOTIFS ProfileScan

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
5	112	T51 T74 S103 S109	N23	DnaJ signature: M7-D79, D11-R82, A22-D41, K30-V97 Nt-DnaJ domains: D26-K42, F59-D79	Protein with DnaJ domain, DNJ1/SIS1 family [Plasmodium falciparum] g3845226	BLAST-GenBank BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS ProfileScan
6	358	T28 S50 T65 S94 S134 T171 T213 S291 S339 Y287 Y303	N63 N194	DnaJ signature: R81-D147, R103-N161 Nt-DnaJ domains: R81-G182 Transmembrane domain: Q231-N251	mDj10 [Mus musculus] g6567172	BLAST-GenBank BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER HMMER-PFAM MOTIFS ProfileScan
7	269	T20 S22 S62 S83 S122 T179	N247	DnaJ domain: E6-E71; F12-S81 Dna J family signatures: A21-D40; E25-K41; F51-E71; R31-K90; C104-F114	DnaJ homolog Hlj1p [S. cerevisiae] g972936	BLAST-Genbank HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS PROFILESCAN BLAST-PRODOM BLAST-DOMO MOTIFS
8	928	S188 T271 S602 S7 T46 T127 S128 S147 T234 S311 S338 S349 T382 S582 S622 S658 S682 T689 T734 S769 T823 S839 S906 S920 S136 T308 S403 T490 T515 T553 S708		DnaJ domain: Y410-F429		MOTIFS

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
9	159	S94 S98 S144 Y95	N10 N156	HSP20 family domain: L37-L125	Chaperone protein [Neisseria meningitidis] g7380239	HMMER-PFAM MOTIFS
10	235	S75 S110 T112 T153 S201 S222 S82 S107 S117	N80 N198	DnaJ domain: D72-L144 HSC domain: F74-K227; D72-I224	Predicted heat shock protein [C. elegans] g1280170	BLAST-Genbank BLAST-PRODOM BLAST-DOMO HMMER-PFAM MOTIFS
11	260	S109 S27 S70 T83 S123 S203 T56 T119 S215 S245		DnaJ domain: A26-D45; D15-G82; G34-R95 Nt-dnaJ domains: E30-R46; F59-D79	Predicted DnaJ protein [C. elegans] g3880170	BLAST-Genbank HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS PROFILESCAN BLAST-PRODOM BLAST-DOMO MOTIFS

Table 3

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
12	172-216 319-363	Reproductive (0.324) Cardiovascular (0.225) Gastrointestinal (0.127)	Inflammation/Trauma (0.408) Cancer (0.352)	PSPORT1
13	94-138 475-519 868-912	Reproductive (0.389) Cardiovascular (0.222) Hematopoietic/Immune (0.222)	Cancer (0.611) Inflammation/Trauma (0.500)	pINCY
14	236-280 608-664	Reproductive (0.275) Nervous (0.176) Gastrointestinal (0.157)	Cancer (0.490) Inflammation/Trauma (0.274) Cell Proliferation (0.176)	PSPORT1
15	388-432	Reproductive (0.344) Nervous (0.188) Gastrointestinal (0.156)	Cancer (0.562) Inflammation/Trauma (0.250) Cell Proliferation (0.188)	PSPORT1
16	329-385	Nervous (0.400) Reproductive (0.400) Urologic (0.200)	Inflammation/Trauma (0.600) Cancer (0.400) Cell Proliferation (0.200)	PBLUESCRIPT
17	196-240 691-735	Gastrointestinal (0.267) Cardiovascular (0.200) Developmental (0.133) Hematopoietic/Immune (0.133)	Cell Proliferation (0.400) Cancer (0.400) Inflammation/Trauma (0.267)	pINCY
18	266-310	Cardiovascular (0.222) Reproductive (0.222) Hematopoietic/Immune (0.167)	Cancer (0.444) Inflammation/Trauma (0.333) Cell Proliferation (0.111)	pINCY
19	462-506	Cardiovascular (0.286) Reproductive (0.179) Hematopoietic/Immune (0.143)	Inflammation/Trauma (0.411) Cancer (0.357) Cell Proliferation (0.125)	pINCY
20	94-138	Reproductive (0.667) Gastrointestinal (0.333)	Cancer (0.667) Inflammation/Trauma (0.333)	pINCY
21	558-602	Reproductive (0.263) Hematopoietic/Immune (0.158) Musculoskeletal (0.158) Nervous (0.158)	Cancer (0.526) Inflammation/Trauma (0.316) Cell Proliferation (0.211)	PSPORT1
22	178-222 515-561	Nervous (0.189) Hematopoietic/Immune (0.162) Reproductive (0.162)	Cancer (0.378) Inflammation/Trauma (0.378) Cell Proliferation (0.378)	pINCY

Table 4

Nucleotide SEQ ID NO:	Library	Library Description
12	SYNOOAT01	Library was constructed using RNA isolated from the knee synovial membrane tissue of an 82-year-old female with osteoarthritis.
13	PROSNOT16	Library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis, and type II diabetes. Family history included benign hypertension, acute myocardial infarcation, hyperlipidemia, and arteriosclerotic coronary artery disease.
14	HIPONON01	This normalized library was constructed from 1.13M independent clones from a hippocampus library. RNA was isolated from the hippocampus tissue of a 72-year-old Caucasian female who died from an intracranial bleed. Patient history included nose cancer, hypertension, and arthritis. The normalization and hybridization conditions were adapted from Soares et al. (Proc. Natl. Acad. Sci. USA (1994) 91:9228).
15	BRSTTUT01	Library was constructed using RNA isolated from breast tumor tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated invasive grade 4 mammary adenocarcinoma of mixed lobular and ductal type, extensively involving the left breast. The tumor was identified in the deep dermis near the lactiferous ducts with extracapsular extension. Seven mid and low and five high axillary lymph nodes were positive for tumor. Proliferative fibrocysytic changes were characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia. Patient history included atrial tachycardia, blood in the stool, and a benign breast neoplasm. Family history included benign hypertension, atherosclerotic coronary artery disease, cerebrovascular disease, and depressive disorder.
16	TESTNOT03	Library was constructed using RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
17	PANCTUT02	Library was constructed using RNA isolated from pancreatic tumor tissue removed from a 45-year-old Caucasian female during radical pancreaticoduodenectomy. Pathology indicated a grade 4 anaplastic carcinoma. Family history included benign hypertension, hyperlipidemia and atherosclerotic coronary artery disease.
18	COLNFET02	This library was constructed using RNA isolated from the colon tissue of a Caucasian female fetus, who died at 20 weeks' gestation.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
19	LSUBNOT03	This library was constructed using RNA isolated from submandibular gland tissue obtained from a 68-year-old Caucasian male during a sialoadenectomy. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
20	OVARNOT07	This library was constructed using RNA isolated from left ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. The tissue was associated with multiple follicular cysts, endometrium in a weakly proliferative phase, and chronic cervicitis with squamous metaplasia. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.
21	LUNGTUT03	This library was constructed using RNA isolated from lung tumor tissue removed from the left lower lobe of a 69-year-old Caucasian male during segmental lung resection. Pathology indicated residual grade 3 invasive squamous cell carcinoma. Patient history included acute myocardial infarction, prostatic hyperplasia, malignant skin neoplasm, and tobacco use.
22	BRAENOT02	This library was constructed using RNA isolated from posterior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	PE Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	PE Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	PE Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-11.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:12-22.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22,

b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-11.

18. A method for treating a disease or condition associated with decreased expression of functional HCPN, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional HCPN, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional HCPN, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and

b) detecting altered expression of the target polynucleotide.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 February 2001 (08.02.2001)

PCT

(10) International Publication Number
WO 01/09178 A3

(51) International Patent Classification[7]: C12N 15/12, C07K 14/47, C12N 15/63, A61K 38/17

(21) International Application Number: PCT/US00/21313

(22) International Filing Date: 3 August 2000 (03.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/146,908 3 August 1999 (03.08.1999) US
60/160,924 22 October 1999 (22.10.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/146,908 (CIP)
Filed on 3 August 1999 (03.08.1999)
US 60/160,924 (CIP)
Filed on 22 October 1999 (22.10.1999)

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). LU, Dyung, Aina, M. [US/US]; 55 Park Belmont Place, San Jose, CA 94304 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report: 27 September 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/09178 A3

(54) Title: HUMAN CHAPERONE PROTEINS

(57) **Abstract:** The invention provides human chaperone proteins (HCPN) and polynucleotides which identify and encode HCPN. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of HCPN.

SEQUENCE LISTING

<110> INCYTE GENOMICS, INC.
YUE, Henry
BANDMAN, Olga
TANG, Y. Tom
BAUGHN, Mariah R.
AZIMZAI, Yalda
LU, Dyung Aina M.

<120> HUMAN CHAPERONE PROTEINS

<130> PF-0728 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/146,908; 60/160,924
<151> 1999-08-03; 1999-10-22

<160> 22
<170> PERL Program

<210> 1
<211> 170
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 723593CD1

<400> 1
Met Ser His Arg Thr Ser Ser Thr Phe Arg Ala Glu Arg Ser Phe
1 5 10 15
His Ser Ser Ser Ser Ser Ser Ser Ser Ser Thr Ser Ser Ser Ala
20 25 30
Ser Arg Ala Leu Pro Ala Gln Asp Pro Pro Met Glu Lys Ala Leu
35 40 45
Ser Met Phe Ser Asp Asp Phe Gly Ser Phe Met Arg Pro His Ser
50 55 60
Glu Pro Leu Ala Phe Pro Ala Arg Pro Gly Gly Ala Gly Asn Ile
65 70 75
Lys Thr Leu Gly Asp Ala Tyr Glu Phe Ala Val Asp Val Arg Asp
80 85 90
Phe Ser Pro Glu Asp Ile Ile Val Thr Thr Ser Asn Asn His Ile
95 100 105
Glu Val Arg Ala Glu Lys Leu Ala Ala Asp Gly Thr Val Met Asn
110 115 120
Thr Phe Ala His Lys Cys Gln Leu Pro Glu Asp Val Asp Pro Thr
125 130 135
Ser Val Thr Ser Ala Leu Arg Glu Asp Gly Ser Leu Thr Ile Arg
140 145 150
Ala Arg Arg His Pro His Thr Glu His Val Gln Gln Thr Phe Arg
155 160 165
Thr Glu Ile Lys Ile
170

<210> 2
<211> 304
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1708350CD1

<400> 2
Met Ala Val Thr Lys Glu Leu Leu Gln Met Asp Leu Tyr Ala Leu
1 5 10 15
Leu Gly Ile Glu Glu Lys Ala Ala Asp Lys Glu Val Lys Lys Ala
20 25 30
Tyr Arg Gln Lys Ala Leu Ser Cys His Pro Asp Lys Asn Pro Asp
35 40 45
Asn Pro Arg Ala Ala Glu Leu Phe His Gln Leu Ser Gln Ala Leu
50 55 60
Glu Val Leu Thr Asp Ala Ala Ala Arg Ala Ala Tyr Asp Lys Val
65 70 75
Arg Lys Ala Lys Lys Gln Ala Ala Glu Arg Thr Gln Lys Leu Asp
80 85 90
Glu Lys Arg Lys Lys Val Lys Leu Asp Leu Glu Ala Arg Glu Arg
95 100 105
Gln Ala Gln Ala Gln Glu Ser Glu Glu Glu Glu Glu Ser Arg Ser
110 115 120
Thr Arg Thr Leu Glu Gln Glu Ile Glu Arg Leu Arg Glu Glu Gly
125 130 135
Ser Arg Gln Leu Glu Glu Gln Gln Arg Leu Ile Arg Glu Gln Ile
140 145 150
Arg Gln Glu Arg Asp Gln Arg Leu Arg Gly Lys Ala Glu Asn Thr
155 160 165
Glu Gly Gln Gly Thr Pro Lys Leu Lys Leu Lys Trp Lys Cys Lys
170 175 180
Lys Glu Asp Glu Ser Lys Gly Gly Tyr Ser Lys Asp Val Leu Leu
185 190 195
Arg Leu Leu Gln Lys Tyr Gly Glu Val Leu Asn Leu Val Leu Ser
200 205 210
Ser Lys Lys Pro Gly Thr Ala Val Val Glu Phe Ala Thr Val Lys
215 220 225
Ala Ala Glu Leu Ala Val Gln Asn Glu Val Gly Leu Val Asp Asn
230 235 240
Pro Leu Lys Ile Ser Trp Leu Glu Gly Gln Pro Gln Asp Ala Val
245 250 255
Gly Arg Ser His Ser Gly Leu Ser Lys Gly Ser Val Leu Ser Glu
260 265 270
Arg Asp Tyr Glu Ser Leu Val Met Met Arg Met Arg Gln Ala Ala
275 280 285
Glu Arg Gln Gln Leu Ile Ala Arg Met Gln Gln Glu Asp Gln Glu
290 295 300
Gly Pro Pro Thr

<210> 3
<211> 483
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1742550CD1

<400> 3
Met Ala Lys Asp Ala Ser Ser Ala Asp Ile Arg Lys Ala Tyr Arg
1 5 10 15
Lys Leu Ser Leu Thr Leu His Pro Asp Lys Asn Lys Asp Glu Asn
20 25 30
Ala Glu Thr Gln Phe Arg Gln Leu Val Ala Ile Tyr Glu Val Leu
35 40 45
Lys Asp Asp Glu Arg Arg Gln Arg Tyr Asp Asp Ile Leu Ile Asn
50 55 60
Gly Leu Pro Asp Trp Arg Gln Pro Val Phe Tyr Tyr Arg Arg Val
65 70 75
Arg Lys Met Ser Asn Ala Glu Leu Ala Leu Leu Leu Phe Ile Ile
80 85 90
Leu Thr Val Gly His Tyr Ala Val Val Trp Ser Ile Tyr Leu Glu
95 100 105

Lys Gln Leu Asp Glu Leu Leu Ser Arg Lys Lys Arg Glu Lys Lys
110 115 120
Lys Lys Thr Gly Ser Lys Ser Val Asp Val Ser Lys Leu Gly Ala
125 130 135
Ser Glu Lys Asn Glu Arg Leu Leu Met Lys Pro Gln Trp His Asp
140 145 150
Leu Leu Pro Cys Lys Leu Gly Ile Trp Phe Cys Leu Thr Leu Lys
155 160 165
Ala Leu Pro His Leu Ile Gln Asp Ala Gly Gln Phe Tyr Ala Lys
170 175 180
Tyr Lys Glu Thr Arg Leu Lys Glu Lys Glu Asp Ala Leu Thr Arg
185 190 195
Thr Glu Leu Glu Thr Leu Gln Lys Gln Lys Lys Val Lys Lys Pro
200 205 210
Lys Pro Glu Phe Pro Val Tyr Thr Pro Leu Glu Thr Thr Tyr Ile
215 220 225
Gln Ser Tyr Asp His Gly Thr Ser Ile Glu Glu Ile Glu Glu Gln
230 235 240
Met Asp Asp Trp Leu Glu Asn Arg Asn Arg Thr Gln Lys Lys Gln
245 250 255
Ala Pro Glu Trp Thr Glu Glu Asp Leu Ser Gln Leu Thr Arg Ser
260 265 270
Met Val Lys Phe Pro Gly Gly Thr Pro Gly Arg Trp Glu Lys Ile
275 280 285
Ala His Glu Leu Gly Arg Ser Val Thr Asp Val Thr Thr Lys Ala
290 295 300
Lys Gln Leu Lys Asp Ser Val Thr Cys Ser Pro Gly Met Val Arg
305 310 315
Leu Ser Glu Leu Lys Ser Thr Val Gln Asn Ser Arg Pro Ile Lys
320 325 330
Thr Ala Thr Thr Leu Pro Asp Asp Met Ile Thr Gln Arg Glu Asp
335 340 345
Ala Glu Gly Val Ala Ala Glu Glu Glu Gln Glu Gly Asp Ser Gly
350 355 360
Glu Gln Glu Thr Gly Ala Thr Asp Ala Arg Pro Arg Arg Arg Lys
365 370 375
Pro Ala Arg Leu Leu Glu Ala Thr Ala Lys Pro Glu Pro Glu Glu
380 385 390
Lys Ser Arg Ala Lys Arg Gln Lys Asp Phe Asp Ile Ala Glu Gln
395 400 405
Asn Glu Ser Ser Asp Glu Glu Ser Leu Arg Lys Glu Arg Ala Arg
410 415 420
Ser Ala Glu Glu Pro Trp Thr Gln Asn Gln Gln Lys Leu Leu Glu
425 430 435
Leu Ala Leu Gln Gln Tyr Pro Arg Gly Ser Ser Asp Arg Trp Asp
440 445 450
Lys Ile Ala Arg Cys Val Pro Ser Lys Ser Lys Glu Asp Cys Ile
455 460 465
Ala Arg Tyr Lys Leu Leu Val Glu Leu Val Gln Lys Lys Lys Gln
470 475 480
Ala Lys Ser

<210> 4
<211> 226
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1919301CD1

<400> 4
Met Ala Ala Met Arg Trp Arg Trp Trp Gln Arg Leu Leu Pro Trp
1 5 10 15
Arg Leu Leu Gln Ala Arg Gly Phe Pro Gln Asn Ser Ala Pro Ser
20 25 30
Leu Gly Leu Gly Ala Arg Thr Tyr Ser Gln Gly Asp Cys Ser Tyr

35 40 45
Ser Arg Thr Ala Leu Tyr Asp Leu Leu Gly Val Pro Ser Thr Ala
50 55 60
Thr Gln Ala Gln Ile Lys Ala Ala Tyr Tyr Arg Gln Cys Phe Leu
65 70 75
Tyr His Pro Asp Arg Asn Ser Gly Ser Ala Glu Ala Ala Glu Arg
80 85 90
Phe Thr Arg Ile Ser Gln Ala Tyr Val Val Leu Gly Ser Ala Thr
95 100 105
Leu Arg Arg Lys Tyr Asp Arg Gly Leu Leu Ser Asp Glu Asp Leu
110 115 120
Arg Gly Pro Gly Val Arg Pro Ser Arg Thr Pro Ala Pro Asp Pro
125 130 135
Gly Ser Pro Arg Thr Pro Pro Pro Thr Ser Arg Thr His Asp Gly
140 145 150
Ser Arg Ala Ser Pro Gly Ala Asn Arg Thr Met Phe Asn Phe Asp
155 160 165
Ala Phe Tyr Gln Ala His Tyr Gly Glu Gln Leu Glu Arg Glu Arg
170 175 180
Arg Leu Arg Ala Arg Arg Glu Ala Leu Arg Lys Arg Gln Glu Tyr
185 190 195
Arg Ser Met Lys Gly Leu Arg Trp Glu Asp Thr Arg Asp Thr Ala
200 205 210
Ala Ile Phe Leu Ile Phe Ser Ile Phe Ile Ile Ile Gly Phe Tyr
215 220 225
Ile

<210> 5
<211> 112
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2012055CD1

<400> 5
Met Met Ala Val Glu Gln Met Pro Lys Lys Asp Trp Tyr Ser Ile
1 5 10 15
Leu Gly Ala Asp Pro Ser Ala Asn Ile Ser Asp Leu Lys Gln Lys
20 25 30
Tyr Gln Lys Leu Ile Leu Met Tyr His Pro Asp Lys Gln Ser Thr
35 40 45
Asp Val Pro Ala Gly Thr Val Glu Glu Cys Val Gln Lys Phe Ile
50 55 60
Glu Ile Asp Gln Ala Trp Lys Ile Leu Gly Asn Glu Glu Thr Lys
65 70 75
Arg Glu Tyr Asp Leu Gln Arg Cys Glu Asp Asp Leu Arg Asn Val
80 85 90
Gly Pro Val Asp Ala Gln Val Tyr Leu Glu Glu Met Ser Trp Asn
95 100 105
Glu Val Thr Ser Gln Arg Gln
110
<210> 6
<211> 358
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2238062CD1

<400> 6
Met Ala Ala Thr Leu Gly Ser Gly Glu Arg Trp Thr Glu Ala Tyr
1 5 10 15
Ile Asp Ala Val Arg Arg Asn Lys Tyr Pro Glu Asp Thr Pro Pro
20 25 30

Glu Ser His Asp Pro Cys Gly Cys Cys Asn Cys Met Lys Ala Gln
35 40 45
Lys Glu Lys Lys Ser Glu Asn Glu Trp Thr Gln Thr Arg Gln Gly
50 55 60
Glu Gly Asn Ser Thr Tyr Ser Glu Glu Gln Leu Leu Gly Val Gln
65 70 75
Arg Ile Lys Lys Cys Arg Asn Tyr Tyr Glu Ile Leu Gly Val Ser
80 85 90
Arg Asp Ala Ser Asp Glu Glu Leu Lys Lys Ala Tyr Arg Lys Leu
95 100 105
Ala Leu Lys Phe His Pro Asp Lys Asn Cys Ala Pro Gly Ala Thr
110 115 120
Asp Ala Phe Lys Ala Ile Gly Asn Ala Phe Ala Val Leu Ser Asn
125 130 135
Pro Asp Lys Arg Leu Arg Tyr Asp Glu Tyr Gly Asp Glu Gln Val
140 145 150
Thr Phe Thr Ala Pro Arg Ala Arg Pro Tyr Asn Tyr Tyr Arg Asp
155 160 165
Phe Glu Ala Asp Ile Thr Pro Glu Glu Leu Phe Asn Val Phe Phe
170 175 180
Gly Gly His Phe Pro Thr Gly Asn Ile His Met Phe Ser Asn Val
185 190 195
Thr Asp Asp Thr Tyr Tyr Tyr Arg Arg Arg His Arg His Glu Arg
200 205 210
Thr Gln Thr Gln Lys Glu Glu Glu Glu Glu Lys Pro Gln Thr Thr
215 220 225
Tyr Ser Ala Phe Ile Gln Leu Leu Pro Val Leu Val Ile Val Ile
230 235 240
Ile Ser Val Ile Thr Gln Leu Leu Ala Thr Asn Pro Pro Tyr Ser
245 250 255
Leu Phe Tyr Lys Ser Thr Leu Gly Tyr Thr Ile Ser Arg Glu Thr
260 265 270
Gln Asn Leu Gln Val Pro Tyr Phe Val Asp Lys Asn Phe Asp Lys
275 280 285
Ala Tyr Arg Gly Ala Ser Leu His Asp Leu Glu Lys Thr Ile Glu
290 295 300
Lys Asp Tyr Ile Asp Tyr Ile Gln Thr Ser Cys Trp Lys Glu Lys
305 310 315
Gln Gln Lys Ser Glu Leu Thr Asn Leu Ala Gly Leu Tyr Arg Asp
320 325 330
Glu Arg Leu Lys Gln Lys Ala Glu Ser Leu Lys Leu Glu Asn Cys
335 340 345
Glu Lys Leu Ser Lys Leu Ile Gly Leu Arg Arg Gly Gly
350 355

<210> 7
<211> 928
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1825012CD1

<400> 7
Met Gly Gly Ser Ala Ser Ser Gln Leu Asp Glu Gly Lys Cys Ala
1 5 10 15
Tyr Ile Arg Gly Lys Thr Glu Ala Ala Ile Lys Asn Phe Ser Pro
20 25 30
Tyr Tyr Ser Arg Gln Tyr Ser Val Ala Phe Cys Asn His Val Arg
35 40 45
Thr Glu Val Glu Gln Gln Arg Asp Leu Thr Ser Gln Phe Leu Lys
50 55 60
Thr Lys Pro Pro Leu Ala Pro Gly Thr Ile Leu Tyr Glu Ala Glu
65 70 75
Leu Ser Gln Phe Ser Glu Asp Ile Lys Lys Trp Lys Glu Arg Tyr
80 85 90
Val Val Val Lys Asn Asp Tyr Ala Val Glu Ser Tyr Glu Asn Lys

```
                    95                  100                 105
Glu Ala Tyr Gln Arg Gly Ala Ala Pro Lys Cys Arg Ile Leu Pro
                110                 115                 120
Ala Gly Gly Lys Val Leu Thr Ser Glu Asp Glu Tyr Asn Leu Leu
                125                 130                 135
Ser Asp Arg His Phe Pro Asp Pro Leu Ala Ser Ser Glu Lys Glu
                140                 145                 150
Asn Thr Gln Pro Phe Val Val Leu Pro Lys Glu Phe Pro Val Tyr
                155                 160                 165
Leu Trp Gln Pro Phe Phe Arg His Gly Tyr Phe Cys Phe His Glu
                170                 175                 180
Ala Ala Asp Gln Lys Arg Phe Ser Ala Leu Leu Ser Asp Cys Val
                185                 190                 195
Arg His Leu Asn His Asp Tyr Met Lys Gln Met Thr Phe Glu Ala
                200                 205                 210
Gln Ala Phe Leu Glu Ala Val Gln Phe Phe Arg Gln Glu Lys Gly
                215                 220                 225
His Tyr Gly Ser Trp Glu Met Ile Thr Gly Asp Glu Ile Gln Ile
                230                 235                 240
Leu Ser Asn Leu Val Met Glu Glu Leu Leu Pro Thr Leu Gln Thr
                245                 250                 255
Asp Leu Leu Pro Lys Met Lys Gly Lys Lys Asn Asp Arg Lys Arg
                260                 265                 270
Thr Trp Leu Gly Leu Leu Glu Glu Ala Tyr Thr Leu Val Gln His
                275                 280                 285
Gln Val Ser Glu Gly Leu Ser Ala Leu Lys Glu Glu Cys Arg Ala
                290                 295                 300
Leu Thr Lys Gly Leu Glu Gly Thr Ile Arg Ser Asp Met Asp Gln
                305                 310                 315
Ile Val Asn Ser Lys Asn Tyr Leu Ile Gly Lys Ile Lys Ala Met
                320                 325                 330
Val Ala Gln Pro Ala Glu Lys Ser Cys Leu Glu Ser Val Gln Pro
                335                 340                 345
Phe Leu Ala Ser Ile Leu Glu Glu Leu Met Gly Pro Val Ser Ser
                350                 355                 360
Gly Phe Ser Glu Val Arg Val Leu Phe Glu Lys Glu Val Asn Glu
                365                 370                 375
Val Ser Gln Asn Phe Gln Thr Thr Lys Asp Ser Val Gln Leu Lys
                380                 385                 390
Glu His Leu Asp Arg Leu Met Asn Leu Pro Leu His Ser Val Lys
                395                 400                 405
Met Glu Pro Cys Tyr Thr Lys Val Asn Leu Leu His Glu Arg Leu
                410                 415                 420
Gln Asp Leu Lys Ser Arg Phe Arg Phe Pro His Ile Asp Leu Val
                425                 430                 435
Val Gln Arg Thr Gln Asn Tyr Met Gln Glu Leu Met Glu Asn Ala
                440                 445                 450
Val Phe Thr Phe Glu Gln Leu Leu Ser Pro His Leu Gln Gly Glu
                455                 460                 465
Ala Ser Lys Thr Ala Val Ala Ile Glu Lys Val Lys Leu Arg Val
                470                 475                 480
Leu Lys Gln Tyr Asp Tyr Asp Ser Ser Thr Ile Arg Lys Lys Ile
                485                 490                 495
Phe Gln Glu Ala Leu Val Gln Ile Thr Leu Pro Thr Val Gln Lys
                500                 505                 510
Ala Leu Ala Ser Thr Cys Lys Pro Glu Leu Gln Lys Tyr Glu Gln
                515                 520                 525
Phe Ile Phe Ala Asp His Thr Asn Met Ile His Val Glu Asn Val
                530                 535                 540
Tyr Glu Glu Ile Leu His Gln Ile Leu Leu Asp Glu Thr Leu Lys
                545                 550                 555
Val Ile Lys Glu Ala Ala Ile Leu Lys Lys His Asn Leu Phe Glu
                560                 565                 570
Asp Asn Met Ala Leu Pro Ser Glu Ser Val Ser Ser Leu Thr Asp
                575                 580                 585
Leu Lys Pro Pro Thr Gly Ser Asn Gln Ala Ser Pro Ala Arg Arg
                590                 595                 600
```

Ala Ser Ala Ile Leu Pro Gly Val Leu Gly Ser Glu Thr Leu Ser
605 610 615
Asn Glu Val Phe Gln Glu Ser Glu Glu Glu Lys Gln Pro Glu Val
620 625 630
Pro Ser Ser Leu Ala Lys Gly Glu Ser Leu Ser Leu Pro Gly Pro
635 640 645
Ser Pro Pro Pro Asp Gly Thr Glu Gln Val Ile Ile Ser Arg Val
650 655 660
Asp Asp Pro Val Val Asn Pro Val Ala Thr Glu Asp Thr Ala Gly
665 670 675
Leu Pro Gly Thr Cys Ser Ser Glu Leu Glu Phe Gly Gly Thr Leu
680 685 690
Glu Asp Glu Glu Pro Ala Gln Glu Glu Pro Glu Pro Ile Thr Ala
695 700 705
Ser Gly Ser Leu Lys Ala Leu Arg Lys Leu Leu Thr Ala Ser Val
710 715 720
Glu Val Pro Val Asp Ser Ala Pro Val Met Glu Glu Asp Thr Asn
725 730 735
Gly Glu Ser His Val Pro Gln Glu Asn Glu Glu Glu Glu Glu Lys
740 745 750
Glu Pro Ser Gln Ala Ala Ala Ile His Pro Asp Asn Cys Glu Glu
755 760 765
Ser Glu Val Ser Glu Arg Glu Ala Gln Pro Pro Cys Pro Glu Ala
770 775 780
His Gly Glu Glu Leu Gly Gly Phe Pro Glu Val Gly Ser Pro Ala
785 790 795
Ser Pro Pro Ala Ser Gly Gly Leu Thr Glu Glu Pro Leu Gly Pro
800 805 810
Met Glu Gly Glu Leu Pro Gly Glu Ala Cys Thr Leu Thr Ala His
815 820 825
Glu Gly Arg Gly Gly Lys Cys Thr Glu Glu Gly Asp Ala Ser Gln
830 835 840
Gln Glu Gly Cys Thr Leu Gly Ser Asp Pro Ile Cys Leu Ser Glu
845 850 855
Ser Gln Val Ser Glu Glu Gln Glu Glu Met Gly Gly Gln Ser Ser
860 865 870
Ala Ala Gln Ala Thr Ala Ser Val Asn Ala Glu Glu Ile Lys Val
875 880 885
Ala Arg Ile His Glu Cys Gln Trp Val Val Glu Asp Ala Pro Asn
890 895 900
Pro Asp Val Leu Leu Ser His Lys Asp Asp Val Lys Glu Gly Glu
905 910 915
Gly Gly Gln Glu Ser Phe Pro Glu Leu Pro Ser Glu Glu
920 925

<210> 8
<211> 159
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1906464CD1

<400> 8
Met Gln Arg Val Gly Asn Thr Phe Ser Asn Glu Ser Arg Val Ala
1 5 10 15
Ser Arg Cys Pro Ser Val Gly Leu Ala Glu Arg Asn Arg Val Ala
20 25 30
Thr Met Pro Val Arg Leu Leu Arg Asp Ser Pro Ala Ala Gln Glu
35 40 45
Asp Asn Asp His Ala Arg Asp Gly Phe Gln Met Lys Leu Asp Ala
50 55 60
His Gly Phe Ala Pro Glu Glu Leu Val Val Gln Val Asp Gly Gln
65 70 75
Trp Leu Met Val Thr Gly Gln Gln Gln Leu Asp Val Arg Asp Pro
80 85 90
Glu Arg Val Ser Tyr Arg Met Ser Gln Lys Val His Arg Lys Met

95 100 105
Leu Pro Ser Asn Leu Ser Pro Thr Ala Met Thr Cys Cys Leu Thr
110 115 120
Pro Ser Gly Gln Leu Trp Val Arg Gly Gln Cys Val Ala Leu Ala
125 130 135
Leu Pro Glu Ala Gln Thr Gly Pro Ser Pro Arg Leu Gly Ser Leu
140 145 150
Gly Ser Lys Ala Ser Asn Leu Thr Arg
155

<210> 9
<211> 235
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1979146CD1

<400> 9
Met Trp Arg Gly Arg Ala Gly Ala Leu Leu Arg Val Trp Gly Phe
1 5 10 15
Trp Pro Thr Gly Val Pro Arg Arg Arg Pro Leu Ser Cys Asp Ala
20 25 30
Ala Ser Gln Ala Gly Ser Asn Tyr Pro Arg Cys Trp Asn Cys Gly
35 40 45
Gly Pro Trp Gly Pro Gly Arg Glu Asp Arg Phe Phe Cys Pro Gln
50 55 60
Cys Arg Ala Leu Gln Ala Pro Asp Pro Thr Arg Asp Tyr Phe Ser
65 70 75
Leu Met Asp Cys Asn Arg Ser Phe Arg Val Asp Thr Ala Asn Val
80 85 90
Gln His Arg Tyr Gln Gln Leu Gln Arg Leu Val His Pro Asp Phe
95 100 105
Phe Ser Gln Arg Ser Gln Thr Glu Lys Asp Phe Ser Glu Lys His
110 115 120
Ser Thr Leu Val Asn Asp Ala Tyr Lys Thr Leu Leu Ala Pro Leu
125 130 135
Ser Arg Gly Leu Tyr Leu Leu Lys Leu His Gly Ile Glu Ile Pro
140 145 150
Glu Arg Thr Asp Tyr Glu Met Asp Arg Gln Phe Leu Ile Glu Ile
155 160 165
Met Glu Ile Asn Glu Lys Leu Ala Glu Ala Glu Ser Glu Ala Ala
170 175 180
Met Lys Glu Ile Glu Ser Ile Val Lys Ala Lys Gln Lys Glu Phe
185 190 195
Thr Asp Asn Val Ser Ser Ala Phe Glu Gln Asp Asp Phe Glu Glu
200 205 210
Ala Lys Glu Ile Leu Thr Lys Met Arg Tyr Phe Ser Asn Ile Glu
215 220 225
Glu Lys Ile Lys Leu Lys Lys Ile Pro Leu
230 235

<210> 10
<211> 260
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5680480CD1

<400> 10
Met Gly Leu Leu Asp Leu Cys Glu Glu Val Phe Gly Thr Ala Asp
1 5 10 15
Leu Tyr Arg Val Leu Gly Val Arg Arg Glu Ala Ser Asp Gly Glu
20 25 30
Val Arg Arg Gly Tyr His Lys Val Ser Leu Gln Val His Pro Asp
35 40 45

Arg Val Gly Glu Gly Asp Lys Glu Asp Ala Thr Arg Arg Phe Gln
50 55 60
Ile Leu Gly Lys Val Tyr Ser Val Leu Ser Asp Arg Glu Gln Arg
65 70 75
Ala Val Tyr Asp Glu Gln Gly Thr Val Asp Glu Asp Ser Pro Val
80 85 90
Leu Thr Gln Asp Arg Asp Trp Glu Ala Tyr Trp Arg Leu Leu Phe
95 100 105
Lys Lys Ile Ser Leu Glu Asp Ile Gln Ala Phe Glu Lys Thr Tyr
110 115 120
Lys Gly Ser Glu Glu Glu Leu Ala Asp Ile Lys Gln Ala Tyr Leu
125 130 135
Asp Phe Lys Gly Asp Met Asp Gln Ile Met Glu Ser Val Leu Cys
140 145 150
Val Gln Tyr Thr Glu Glu Pro Arg Ile Arg Asn Ile Ile Gln Gln
155 160 165
Ala Ile Asp Ala Gly Glu Val Pro Ser Tyr Asn Ala Phe Val Lys
170 175 180
Glu Ser Lys Gln Lys Met Asn Ala Arg Lys Arg Arg Ala Gln Glu
185 190 195
Glu Ala Lys Glu Ala Glu Met Ser Arg Lys Glu Leu Gly Leu Asp
200 205 210
Glu Gly Val Asp Ser Leu Lys Ala Ala Ile Gln Ser Arg Gln Lys
215 220 225
Asp Arg Gln Lys Glu Met Asp Asn Phe Leu Ala Gln Met Glu Ala
230 235 240
Lys Tyr Cys Lys Ser Ser Lys Gly Gly Gly Lys Lys Ser Ala Leu
245 250 255
Lys Lys Glu Lys Lys
260

<210> 11
<211> 269
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1459372CD1

<400> 11
Met Ala Gly Val Pro Glu Asp Glu Leu Asn Pro Phe His Val Leu
1 5 10 15
Gly Val Glu Ala Thr Ala Ser Asp Val Glu Leu Lys Lys Ala Tyr
20 25 30
Arg Gln Leu Ala Val Met Val His Pro Asp Lys Asn His His Pro
35 40 45
Arg Ala Glu Glu Ala Phe Lys Val Leu Arg Ala Ala Trp Asp Ile
50 55 60
Val Ser Asn Ala Glu Lys Arg Lys Glu Tyr Glu Met Lys Arg Met
65 70 75
Ala Glu Asn Glu Leu Ser Arg Ser Val Asn Glu Phe Leu Ser Lys
80 85 90
Leu Gln Asp Asp Leu Lys Glu Ala Met Asn Thr Met Met Cys Ser
95 100 105
Arg Cys Gln Gly Lys His Arg Arg Phe Glu Met Asp Arg Glu Pro
110 115 120
Lys Ser Ala Arg Tyr Cys Ala Glu Cys Asn Arg Leu His Pro Ala
125 130 135
Glu Glu Gly Asp Phe Trp Ala Glu Ser Ser Met Leu Gly Leu Lys
140 145 150
Ile Thr Tyr Phe Ala Leu Met Asp Gly Lys Val Tyr Asp Ile Thr
155 160 165
Glu Trp Ala Gly Cys Gln Arg Val Gly Ile Ser Pro Asp Thr His
170 175 180
Arg Val Pro Tyr His Ile Ser Phe Gly Ser Arg Ile Pro Gly Thr
185 190 195
Arg Gly Arg Gln Arg Ala Thr Pro Asp Ala Pro Pro Ala Asp Leu

                200   205   210
Gln Asp Phe Leu Ser Arg Ile Phe Gln Val Pro Pro Gly Gln Met
                215   220   225
Pro Asn Gly Asn Phe Phe Ala Ala Pro Gln Pro Ala Pro Gly Ala
                230   235   240
Ala Ala Ala Ser Lys Pro Asn Ser Thr Val Pro Lys Gly Glu Ala
                245   250   255
Lys Pro Lys Arg Arg Lys Lys Val Arg Arg Pro Phe Gln Arg
                260   265

<210> 12
<211> 1550
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 723593CB1

<400> 12
gtcggagcct ggcacgctcg cccagaggcc tgcgcccaca ccctctcctg tccagccctc 60
gcccgcctgg gcagggcccg gcgccgtccg tggatgagcc acagaacctc ttccaccttc 120
cgagcggaga gaagtttcca ttcctcttct tcttcctcct cctcttccac ctcctcctcg 180
gcctcccgtg ccctcccggc ccaggacccg cccatggaga aggccctgag catgttttcc 240
gatgactttg gcagcttcat gcggccccac tcggagcccc tggccttccc agcccgcccc 300
ggtggggcag gcaacatcaa gaccctagga gacgcctatg agtttgcggt ggacgtgaga 360
gacttctcac ctgaagacat cattgtcacc acctccaaca accacatcga ggtgcgggct 420
gagaagctgg cggctgacgg cactgtcatg aacaccttcg ctcacaagtg ccagctgccg 480
gaggacgtgg acccgacgtc ggtgacctcg gctctgcggg aggacggcag cctcactatc 540
cgggcacggc gtcacccgca tacagaacac gtccagcaga ccttccggac ggagatcaaa 600
atctgagtgc ctctcccttc cctttccctg tcccccgcc ccacgcctgc cagcaaagcc 660
tcgctaaccc cattacaaca gctccaggac atctcagccc aggttctagc cccacgcac 720
cccagacccc aggtggacca tcctcccaaa ctagggccct ccactctatc cagggcaggc 780
cagggactcc ctggcctgac acatgatgcc cagatttcag atttggcctc cgtcacttaa 840
tccagagtac aggggctggg gtcagggaag gaagatctaa agaacccact gtgggtcagg 900
ggaatgggac cagcaggaca tatgggcaag ctctgcagga cagacagaca gacaaaccct 960
ctgatctatg aagtctctgc agggcaaggg gaccagggac ctggaaccct cttggccaag 1020
gggagtggga gggacagagg gaaggtcaca ggcaagggtg cctatctaag tggaactaat 1080
tgcccgaggg ctcagcaagg ccaagaggag acagccgtga cggtaaactt cccctctacc 1140
agcctccaag ccccacgcca gcgagcaggc tgcctgccca ccccgtgccc ccagccagct 1200
ggctgtgcca gggcagagcc atgccacatc tgtatataga tggggttttt ccaatacagc 1260
tggttcgtga taaactgcat gaaactcctg ccgtcctgcg cctgctgggg cctccaggca 1320
aggccacgtg ggttgggggg tggggctggt ccttctccct cccacaggcc tgtgttcttg 1380
gggctgctcc catgcagaca ggatcaccta acagagatgg aagccagggc atggatgggg 1440
ctttgggtcc tcaaggttgg accccagctt cttgccacct tcccctccgg gcagtcagct 1500
ctccatccat ccccctcttt aatctatgaa tctataggct cggtgtgtgt 1550

<210> 13
<211> 1075
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1708350CB1

<400> 13
cagaacacaa ttcccagagg gctaggcgcc gctcggagcc tgcagtcctc acgcgcgctt 60
agactcttgg gagttgtagt acgaatccgt caggccggaa ccatggcagt gaccaaggag 120
ctcttacaga tggacctgta cgcgctgcta ggcattgagg agaaggcagc ggacaaagag 180
gtaaagaagg cgtataggca gaaggccctc tcctgccacc cagacaaaaa tccagataat 240
cccagagcag ctgaactctt ccaccagctt tctcaggcct tggaggtgct gaccgatgct 300
gcagccaggg ctgcatatga caaggtcagg aaagccaaga agcaagcagc agagaggacc 360
cagaaacttg atgagaaaag gaagaaagtg aagcttgacc tggaggcccg gagcggcag 420
gcccaggccc aggagagtga ggaggaagag gagagccgga gcaccaggac actagagcaa 480
gagatcgaac gcctgagaga agagggttcc cggcagctgg aggaacagca gaggctcatc 540
cgggagcaga tacgccagga gcgtgaccag aggttgagag gaaaggcaga aaatactgaa 600
ggccaaggaa ccccccaaact aaagctaaaa tggaagtgca agaaggagga tgagtcaaaa 660

```
ggtggctact ccaaagacgt cctcctacgg cttttgcaga agtatggtga ggttctcaac 720
ctggtgcttt ccagtaagaa gccaggcact gctgtggtgg agtttgcaac cgtcaaggca 780
gcggagctgg ctgtccagaa tgaagttggc ctggtggata accctctgaa gatttcctgg 840
ttggagggac agccccagga tgccgtgggc cgcagccact caggactgtc aaagggctca 900
gtgctgtcag agagggacta cgagagcctc gtcatgatgc gcatgcgcca ggcggccgag 960
cggcaacagc tgatcgcacg gatgcagcag gaagaccagg aggggccgcc tacgtagccc 1020
cagctccagc catccacccg tcagcccttt tcttcaacgt cgcgagataa attta      1075
```

<210> 14
<211> 1950
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1742550CB1

<400> 14

```
aagagagccc gaggcaggta gctgcagttc ccctccaaga cttctccaca cctgtttgac 60
caggtacaag atcaggcgcc ggggtcatct gttcactagg ccacggggtc aggacaagag 120
tcacccgcag ctctgaggcc agatggtaat tccaatcgcc tccccagttc agcagcgaac 180
ccagcaagac gaagataatt ttcgaaacat tcaggctcgg gagtagacgt cgcaatggag 240
tgctgtcctc gcggctttgg agccacgggg catggccaag gatgcatcat ctgcagacat 300
cagaaaagca tatcgtaagc tttcactaac tttacatcca gacaagaata aagatgaaaa 360
tgcagaaact cagtttagac aattggtggc catttatgaa gttttaaagg atgatgaacg 420
aaggcagagg tatgatgata ttctgatcaa tggacttcca gattggcgac agcctgtatt 480
ctactacagg cgggtgagaa aaatgagcaa tgctgagctg gcattactct tgttcattat 540
tctcacagtg ggtcattatg ctgtggtttg gtcaatctac ctggaaaaac aactggatga 600
actactaagt agaaaaaaga gagaaaagaa aaaaaagact ggcagcaaga gtgtggatgt 660
atcaaaactc ggtgcttcag aaaaaaatga aagattgctg atgaaaccac agtggcatga 720
tttgcttcca tgcaaactgg ggatttggtt ttgccttaca ctaaaagcat tacctcacct 780
catccaggat gctgggcagt tttatgctaa atataaagaa acaagattga aggaaaagga 840
agatgcactg actagaactg aacttgaaac acttcaaaaa cagaagaaag ttaaaaaacc 900
aaaacctgaa tttcctgtat acacaccttt agaaactaca tatattcagt cttatgatca 960
tggaacttcc atagaagaaa ttgaggaaca aatggatgat tggttggaaa acaggaaccg 1020
aacacagaaa aaacaggcac ctgaatggac agaagaggac ctcagccaac tgacaagaag 1080
tatggttaag ttcccaggag ggactccagg tcgatgggaa aagattgccc acgaattggg 1140
tcgatctgtg acagatgtga caaccaaagc caagcaactg aaggattcag tgacctgctc 1200
cccaggaatg gttagactct ccgaactcaa atcgacagtt cagaattcca ggcccatcaa 1260
aacggccacc accttgcccg atgacatgat cacccagcga gaggacgcag agggggtggc 1320
agcggaggag gagcaggagg gagactccgg tgagcaggag accgggggca ctgatgcccg 1380
gcctcggagg cggaagccag ccaggctgct ggaggctaca gcgaagccgg agccagagga 1440
gaagtccaga gccaagcggc agaaggactt tgacatagca gaacaaaacg agtccagcga 1500
cgaggagagc ctgagaaaag agagagctcg gtctgcagag gagccgtgga ctcaaaatca 1560
acagaaactt ctggaactgg cgttgcagca gtacccaagg ggatcctctg accgctggga 1620
caaaatagcc agatgtgtcc cgtccaagag caaggaagac tgtatcgcta ggtacaagtt 1680
gctggttgaa ctggtccaaa agaaaaaaca agctaaaagc tgaatattct gggagatgat 1740
gttcaccttc atttccaaa atgaatatct taaaaatctt atgcagaaat ttgcattttg 1800
tacctcaata tttctacgtc atgtgcctta gtaaaaaaaa ataataaata aataaaagat 1860
gagtgttgtg ctaaaaaaaa aaaaaaaaaa aaaaaactcg gtcgcaagct tattcccttt 1920
agtgagggtt aattttagct tgcactggcc                                  1950
```

<210> 15
<211> 1187
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1919301CB1

<400> 15

```
ctcttgcacc gcctgccgaa tcaattcaac atggcagcca tgcgctggcg atggtggcag 60
cggctgttac cttggaggtt gctgcaggcc cgtggctttc cacaaaattc tgcacccagc 120
ctgggcctag gagcgaggac ttattcccag ggcgactgct cgtattcgcg cacggcgctg 180
tatgatctgc tcggcgtccc ctccacagcc acgcaggccc aaatcaaggc ggcttactac 240
cgtcagtgct ttctctacca cccggaccgc aactccggga gcgcggaggc cgccgagcgc 300
```

ttcacgcgca tctcccaggc ctacgtggtg ctgggcagtg ccaccctccg tcgcaagtat 360
gatcgcggcc tactcagcga cgaggacctg cgcggacctg gcgtccggcc ctccaggacg 420
cccgcacccg accccggctc gccgcgtacc ccgccgccca cctctcggac ccacgacggt 480
tctcgggcct cccccggcgc caaccgcacg atgttcaact ttgacgcctt ctaccaggcc 540
cactatgggg aacaactgga gcgggaacgg cgcctgaggg cccggcggga ggcccttcgc 600
aaacggcagg agtatcggtc catgaaaggc ctccgctggg aggatacccg agacacggct 660
gccattttcc tcatcttttc aatcttcatc atcatcggct tttatattta atcggagaga 720
gaagggaagg ggagtgtccc cagccaaccc cccagaaacg gcctttttc ctgcctctga 780
acccttggcc gttgatagtc taccttttgct gggatccgaa ggaactgtac tccccctgcc 840
ctccccgacc cgcccagctt agccgatgac ctgcacatcg ctccactgtg gtccagaaaa 900
ggaggccttt cgatgtctga gaaagaggcc ccacgctgta gagtcccgaa agcccaggag 960
tgaagggggt tcctggagtc tctagggtgc ttcttccaga gtctgtcttc ttgcttccag 1020
atgtggtcaa cttctggaac actcgctgta gctttattgt ttagccccaa gcaagattta 1080
tctcctcctg ccccgcatgt gtatggtggg cctctgtaac cttgaaatgt gcaatgtgac 1140
caattgttga ctaccaaaag aaaaggtctg gggttgtaaa aaaaaaa 1187

<210> 16
<211> 740
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2012055CB1

<400> 16
cgaggagtgg gtagcagcgc ctatgtgaag ttagctaatc tgagaaggcc cacttctggt 60
tccatggatg atggcggttg agcagatgcc aaaaaaggat tggtacagca tcctgggagc 120
agacccatct gcaaatatat cagacctaaa acaaaaatat caaaaactca tattaatgta 180
tcatccagat aaacaaagta cagatgtacc agcaggaaca gtggaggaat gtgtacagaa 240
gttcatcgaa attgatcaag catggaaaat tctaggaaat gaagagacaa aaagagagta 300
tgacctgcag cggtgtgaag atgatctaag aaatgtagga ccagtagatg ctcaagtata 360
tcttgaagaa atgtcttgga atgaagttac ttctcagaga cagtaaaatg gaatgaccaa 420
tggatcagag attctttaag tcaaagggca caagcatttc aacttcccag gaaaatgaca 480
cacttaaaat ttccacgatc aggagcctaa gtattgcacc gtattgcctc ctttgggcat 540
ctcacttcag catcttgttg gttcatgtat catttgtaaa catcaaacac acacacacat 600
acccccatag atttaaaaaa acaacaacaa catggtgttg tgtttataga cttaagtcaa 660
gattcttgaa atagtgtgac actagaagag aaagtatcca gatgttgcat ttgataaata 720
gtctggcttt ctctaaagga 740

<210> 17
<211> 1361
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2238062CB1

<400> 17
tcgggcgcgg gggaggctcg gcggacctgc tgattgggaa ccgatatggc ggcgactctg 60
ggcagcgggg agcgctggac ggaagcttac attgacgcag ttagaagaaa caaataccca 120
gaagacacac ctcctgagag tcatgacccc tgtggctgct gtaactgcat gaaggcacaa 180
aaggaaaaga agtctgagaa tgagtggact cagacccggc agggtgaggg gaactccacg 240
tacagtgagg aacagctgct tggggtacaa aggatcaaga aatgcagaaa ttactatgaa 300
attctgggag tttctcgaga tgctagtgac gaagagctta agaaagctta cagaaaactc 360
gccctgaaat ttcaccctga caagaactgt gctcctggag caacagatgc tttcaaagca 420
ataggaaatg catttgcagt cctgagcaat cctgataaga gacttcgcta tgatgaatac 480
ggagatgaac aggtgacttt cactgcccct cgagccagac cttataatta ttacagggat 540
tttgaagctg acatcactcc agaagagctg ttcaacgtct tctttggagg acattttcct 600
acaggaaata ttcatatgtt ttcaaatgtg acagatgaca cttactatta ccgtcgacgg 660
caccgacatg agaggacaca gactcagaag gaggaggaag aagagaaacc tcagactaca 720
tattctgcat ttattcagct acttccagtt cttgtgattg tgattatatc tgtcattact 780
cagctgctgg ctactaatcc cccatatagt ctgttctata aatcgacctt gggctacacc 840
atttctagag aaactcagaa cctgcaggtg ccttactttg tggataaaaa ctttgacaag 900
gcctacagag gagcttctct gcatgacttg gagaaaacaa tagagaagga ttacattgat 960
tatatccaga ctagttgttg gaaggagaaa caacaaaagt cagagctgac aaatttggca 1020

ggattataca gagatgaacg attgaaacag aaagcagagt cgctgaaact tgaaaactgt 1080
gagaaacttt ccaaactcat tggcctacgc agaggtggct gagaggataa tggtcctacg 1140
cagggctggg gttttgctac ttgttcctat ttatgttcct gattccattt tataatacaa 1200
aactaggtaa tgatgaacac tttactattt gctaacttcg ttggttgggc agagtggcag 1260
gagcatgggc acgagagcca gatgtgtctt cacaggatcc ttcctgggga gtggctccag 1320
ggaccaggag tagttcatct aagttaaatt aatggcaagg c 1361

<210> 18
<211> 4475
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1825012CB1

<400> 18
cgcctctcga aggaagtttg ctcttaattt cagagccggg ttcgccgtcg gatcaacctc 60
caggagctag cagcgggcgc ggaccgggca gtttccgcgc tcagcacagg cagctcgcgg 120
tcatgggcgg ctcagcctcc agccagctgg acgagggcaa gtgcgcttac atccgaggga 180
aaactgaggc tgccatcaaa aacttcagtc cctactacag tcgtcagtac tctgtggctt 240
tctgcaatca cgtgcgcact gaagtagaac agcaaagaga tttaacgtca cagtttttga 300
agaccaagcc accattggcg cctggaacta ttttgtatga agcagagcta tcacaatttt 360
ctgaagacat aaagaagtgg aaggagagat acgttgtagt taaaaatgat tatgctgtgg 420
agagctatga gaataaagag gcctatcaga gaggagctgc tcctaaatgt cgaattcttc 480
cagccggtgg caaggtgtta acctcagaag atgaatataa tctgttgtct gacaggcatt 540
tcccagaccc tcttgcctcc agtgagaagg agaacactca gccctttgtg gtcctgccca 600
aggaattccc agtgtacctg tggcagccct tcttcagaca cggctacttc tgcttccacg 660
aggctgctga ccagaagagg tttagtgccc tcctgagtga ctgcgtcagg catctcaatc 720
atgattacat gaagcagatg acatttgaag cccaagcctt tttagaagct gtgcaattct 780
tccgacagga gaagggtcac tatggttcct gggaaatgat cactggggat gaaatccaga 840
tcctgagtaa cctggtgatg gaggagctcc tgcccactct tcagacagac ctgctgccta 900
agatgaaggg gaagaagaat gacagaaaga ggacgtggct tggtctcctc gaggaggcct 960
acaccctggt tcagcatcaa gtttcagaag gattaagtgc cttgaaggag gaatgcagag 1020
ctctgacaaa gggcctggaa ggaacgatcc gttctgacat ggatcagatt gtgaactcaa 1080
agaactattt aattggaaag atcaaagcga tggtggccca gccggcggag aaaagctgct 1140
tggagagtgt gcagccattc ctggcatcca tcctggagga gctcatggga ccagtgagct 1200
cgggattcag tgaagtacgt gtactctttg agaaagaggt gaatgaagtc agccagaact 1260
tccagaccac caaagacagt gtccagctaa aggagcatct agaccggctt atgaatcttc 1320
cgctgcattc cgtgaagatg gaaccttgtt atactaaagt caacctgctt cacgagcgcc 1380
tgcaggatct caagagccgc ttcagattcc cccacattga tctggtggtt cagaggacac 1440
agaactacat gcaggagcta atggagaatg cagtgttcac ttttgagcag ttgctttccc 1500
cacatctcca aggagaggcc tccaaaactg cagttgccat tgagaaggtt aaactccgag 1560
tcttaaagca atatgattat gacagcagca ccatccgaaa gaagatattt caagaggcac 1620
tagttcaaat cacacttccc actgtgcaga aggcactggc gtccacatgc aaaccagagc 1680
ttcagaaata cgagcagttc atctttgcag atcataccaa tatgattcac gttgaaaatg 1740
tctatgagga gattttacat cagatcctgc ttgatgaaac tctgaaagtg ataaaggaag 1800
ctgctatctt gaagaaacac aacttatttg aagataacat ggccttgccc agtgaaagtg 1860
tgtccagctt aacagatcta aagcccccca cagggtcaaa ccaggccagc cctgccagga 1920
gagcttctgc cattctgcca ggagttctgg gtagtgagac cctcagtaac gaagtattcc 1980
aggagtcaga ggaagagaag cagcctgagg tccctagctc gttggccaaa ggagaaagcc 2040
tttctctccc tggcccaagc ccaccccag atgggactga gcaggtgatt atttcaagag 2100
tggatgaccc cgtggtgaat cctgtggcaa cagaggacac agcaggactc ccgggcacat 2160
gctcatcaga gctggagttt ggagggaccc ttgaggatga agaacccgcc caggaagagc 2220
cagaacccat cactgcctcg ggttctttga aggcgctcag aaagttgctg acagcgtccg 2280
tggaagtacc agtggactct gctccagtga tggaagaaga tacgaatggg gagagccacg 2340
ttccccaaga aaatgaagaa gaagaggaaa aagagcccag tcaggcagct gccatccacc 2400
ccgacaactg tgaagaaagt gaagtcagcg agagggaggc ccaacctccc tgtcccgagg 2460
cccatgggga ggagttgggg ggatttccag aggtaggcag cccagcctct ccgccagcca 2520
gtggagggct caccgaggag cccctggggc ccatggaggg ggagctccca ggagaggcct 2580
gcacactcac tgcccatgaa ggaagagggg gcaagtgtac cgaggaaggg gatgcctcac 2640
agcaagaggg ctgcacctta ggttctgacc ccatctgcct cagtgagagc caggtttctg 2700
aggaacaaga agagatggga gggcaaagca gcgcggccca ggccacggcc agtgtgaatg 2760
cagaggagat caaggtagcc cgtattcatg agtgtcagtg ggtggtggag gatgctccaa 2820
acccggatgt cctgctgtca cacaaagatg acgtgaagga gggagaaggt ggtcaggaga 2880
gtttcccaga gctgccctca gaggagtgaa agggacaatt tggctgaagt ctttctctga 2940
aaaaagccaa agggttatag gggtacactt aggggttgca tgcaagctgt taccaaaaaa 3000

tttttaagta ttttcttaat ttgaataata aaaccagagg aaatgcatac agggcatgag 3060
caactgaggc aaacctttgt ggacatgaat tgttctacga tgaatttttg ctttagtatt 3120
ttaataagaa ttacaaagac aatggcatac ttggggtgag agggagctga ggatgtctga 3180
ggagggaata gtattgcagg gaagactgag aaaacagtag gatgacagtt ttgagtatac 3240
tctgcacttt tcaattgtgc aatcttcttg tgcactttaa ggctttttaa ttttgtttga 3300
gaatgcaaat gtatactgta agtctacctt tactatctac tatgcctact tcaccatctc 3360
ttaaggactc ggcatttgtc cacagtcaga ctgcaagaga gggtaggtca tgaacagtca 3420
cccatgctgg ctgtagcccc cacagaggca atcatgccca atagattcaa gagaagctaa 3480
gcggaaatgg agggcggaag gtgtgatctg tgggactgtc tgggcctgtt actcatcctg 3540
ctatcaattt cttattaatt aatcttgatg attcttatta attaatcaca tttgcaggaa 3600
attcagatga ggcaagaaaa ttttattggc ctgggtaaga ctgaaagcat tccaaattag 3660
gcttagactg tgcaaagggc ttagctaagt tatcgagctt aaaacccgtc aattaaacaa 3720
acattatttg aacagttact gcatgccacg cactgtgttg ggcttagtaa taaaaaaaag 3780
aaaagataag tgcttgttct agcataaatt aaaaggtcca agggaattta atctggaaga 3840
gaacatatgc caatttttaa actatgacag cttttttttt tctctttcca ttcaaatagt 3900
cctggttcat tcccagaagg gcacaaaatg aatgaataaa taaataaatg aataaagaca 3960
aaagccaagg tgtatgctct caagttccaa agatgttatc aaaagctgaa atcatttgtt 4020
tggtcattca gcaagctaat tgagtctctg ttatatacca agcactgggg ataccatggc 4080
gaaaaacaac tttgttcctt cctcctagaa cttacatttt aatggaaata gacaaaacac 4140
atcttcttaa cggatggtga cctataacca ttaatgttga aaatggaaga gacttgcttc 4200
caaaagatta aaaggagttg ttcttttctc cttcagaaaa ataccagatc atttcctaaa 4260
atctccagtc ccaagtatta catcgtggtt tccctccccg actttttatt ttattttatt 4320
ctattttttt gagatggagt ctcactctgt cgccaaggct ggagtgcagt ggtgtgatct 4380
cggctcactg caacctccgc ctcctgggtt caggagtttc tcctctgtca gcgtcccaag 4440
tagctggaat tacagccatg cggcaccatt cccgg 4475

<210> 19
<211> 636
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1906464CB1

<400> 19
gaaccgagcg agcggagctg agctcgggta ggccgcgcga ggtccctcct ctccgggcgt 60
ccgtgcgcct agctctgcgc tgggagcctc gcgccctttg acagcagtta gttgctgact 120
cggatgcaga gagtcggtaa caccttctcc aacgagagcc gggtggcatc ccggtgtccc 180
agcgtgggcc ttgctgaacg gaaccgggtg gccacaatgc cggtgcggct gctcagggac 240
agtccagcgg ctcaggagga caatgaccat gccagagacg gtttccaaat gaagctggat 300
gcccacggct tcgccccgga ggaactggtg gtgcaggtgg atggccaatg gctgatggtg 360
accggacagc agcaactgga cgtcagggac ccggaaaggg tcagttaccg catgtcacag 420
aaggtgcacc ggaaaatgct cccgtccaac ctgagtccta ccgccatgac ctgctgcctg 480
acccccctccg ggcagctgtg ggtcagaggc cagtgtgtgg cgctggccct ccctgaagcc 540
caaacaggac cgtccccgag actcgggagc ctcggctcta aggcttccaa cctgacccgg 600
taaacaaacg acgcgatgtg cagcaaaaaa aaaaaa 636

<210> 20
<211> 1090
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1979146CB1

<400> 20
gcttttcccc acgagtgacc acggctagat aggccgccgg ccagatgtgg cgggggagag 60
ccggggcttt gctccgggtg tgggggtttt ggccgacagg ggttcccaga aggagaccgc 120
taagctgcga tgctgcgtcg caggcgggaa gcaattatcc ccgctgttgg aactgcggcg 180
gcccatgggg ccccgggcgg gaggacaggt tcttctgccc acagtgccga gcgctgcagg 240
cacctgaccc cactcgagac tacttcagcc ttatggactg caaccgttcc ttcagagttg 300
atacagcgaa cgtccagcac aggtaccagc aactgcagcg tcttgtccac ccagatttct 360
tcagccagag gtctcagact gaaaaggact tctcagagaa gcattcgacc ctggtgaatg 420
atgcctataa gaccctcctg gccccccctga gcagaggact gtaccttcta aagctccatg 480
gaatagagat tcctgaaagg acagattatg aaatggacag gcaattcctc atagaaataa 540

tggaaatcaa tgaaaaactc gcagaagctg aaagtgaagc tgccatgaaa gagattgaat 600
ccattgtcaa agctaaacag aaagaattta ctgacaatgt gagcagtgct tttgaacaag 660
atgactttga agaagccaag gaaattttga caaagatgag atacttttca aatatagaag 720
aaaagatcaa gttaaagaag attccccttt aattgtggat agtttaaagt ttaaaaaata 780
aagttcttgc tgggcacagt ggctcacacc tgtaatccca gcactttggg aggctgaggt 840
gggtggatga caaggtcagg agttcaagac cagcttggcc aacatagtga aaccccgtct 900
ctgctgaaaa tacaaaaatt agccgggcat ggtggcgcgt gcctgtaatc ccagctactt 960
ggtaggccga ggcaggagaa tcgcttaaac ccgtgaggtg gaggttgcag tgagcagaga 1020
tcacgcaact gcactccagc ttgggcaaca gagtgagctt aatcttgaaa aataaataaa 1080
tgaaaatgat 1090

<210> 21
<211> 1447
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5680480CB1

<400> 21
cgaaaaagaa gcagtcctgg gttgtacccg gcgcacgtgg gagcggctgc ttcctccggg 60
gtcgtatctc cgcccggcat ggggctgctg gacctttgcg aggaagtgtt cggcaccgcc 120
gacctttacc gggtgctggg cgtgcgacgc gaggcctccg acggcgaggt ccgacgaggc 180
taccacaagg tgtccctgca ggtacacccg gaccgggtgg gtgagggcga caaggaggac 240
gccacccgcc gcttccagat cctgggaaaa gtctattccg ttctcagtga cagagaacag 300
agagcagtgt acgatgagca gggaacagtg gacgaggact ctcctgtgct cacccaagac 360
cgagactggg aggcgtattg gcggctactc tttaaaaaga tatctttaga ggacattcaa 420
gcttttgaaa agacatacaa aggttcggaa gaagagctgg ctgatattaa gcaggcctat 480
ctggacttca agggtgacat ggatcagatc atggagtctg tgctttgcgt gcagtacaca 540
gaggaaccca ggataaggaa tatcattcag caagctattg acgccggaga ggtcccatcc 600
tataatgcct ttgtcaaaga atcgaaacaa aagatgaatg caaggaaaag gagggctcag 660
gaagaggcca aagaagcaga aatgagcaga aaggagttgg ggcttgatga aggcgtggat 720
agcctgaagg cagccattca gagcagacaa aaggatcggc aaaaggaaat ggacaatttt 780
ctggctcaga tggaagcaaa gtactgcaaa tcttccaaag gaggagggaa aaaatctgct 840
ctcaagaaag aaaagaaata atggaatttt tctcttcaaa ggtccttagg tgtaaattga 900
tgccatcgta ggcaaggtgc aggcaggatt tgaaggcaaa agtcaattca gctcttgaga 960
aaaggtgtct ttccagcctg aatttttcag attgactaga ccaagcagaa tctctcaacc 1020
tgatcttagt atttcctaga aagcacttga cattgtgtga ggtctcacct gaaggaactt 1080
ggtggtgaca tttgggaggg tggagggagg cagtgtcctt cctgacagca cttgcctcca 1140
tggatcttct gtacacagaa ctcttatcta ggatgtggtt ctgttcatgc tgctttctgc 1200
gatgtgcgtg tctgttagaa taggctctct acccagctag aacaccttcc agacacttgc 1260
tggacagcta tcttccacat acttcccagt ttacatttgg tcttaatgat cttgaataga 1320
tcctctcttc attttactca gccaggtttt gtactgatgt acaggtgtta aattacttca 1380
agcatttttg taagaggtgt atataattca ataaaaaagg taaaacatga tgattaaaaa 1440
aaaaaaa 1447

<210> 22
<211> 1147
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1459372CB1

<400> 22
gccttgggtc aagcagaata ttaataggca ggggaatgca cctgtagcta gtgggcgcta 60
ctgccagcct gaagaggaag tggctcgact cttgaccatg gctggggttc ctgaggatga 120
gctaaaccct ttccatgtac tgggggttga ggccacagca tcagatgttg aactgaagaa 180
ggcctataga cagctggcag tgatggttca tcctgacaaa aatcatcatc cccgggctga 240
ggaggccttc aaggttttgc gagcagcttg ggacattgtc agcaatgctg aaaagcgaaa 300
ggagtatgag atgaaacgaa tggcagagaa tgagctgagc cggtcagtaa atgagtttct 360
gtccaagctg caagatgacc tcaaggaggc aatgaatact atgatgtgta gccgatgcca 420
aggaaagcat aggaggtttg aaatggaccg ggaacctaag agtgccagat actgtgctga 480
gtgtaatagg ctgcatcctg ctgaggaagg agacttttgg gcagagtcaa gcatgttggg 540
cctcaagatc acctactttg cactgatgga tggaaaggtg tatgacatca cagagtgggc 600

```
tggatgccag cgtgtaggta tctccccaga tacccacaga gtcccctatc acatctcatt 660
tggttctcgg attccaggca ccagagggcg gcagagagcc accccagatg cccctcctgc 720
tgatcttcag gatttcttga gtcggatctt tcaagtaccc ccagggcaga tgcccaatgg 780
gaacttcttt gcagctcctc agcctgcccc tggagccgct gcagcctcta agcccaacag 840
cacagtaccc aagggagaag ccaaacctaa gcggcggaag aaagtgagga ggcccttcca 900
acgttgatgc cccttctctt tcctcaaatc aatgtcaggg agtcaaaagg gctgtagcac 960
aggatggagt ttgatttatc cctcctcccc caacacctag gaactgaatc tttttctttt 1020
tatttttga gatggagtct tgctctgttg cccagctgga gtgcagtggt gtgatctcag 1080
cttactgcaa cctctgtctc ccgggttcaa gcaattctcc catctcagcc tcctgagtag 1140
ctgggat                                                           1147
```

DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

HUMAN CHAPERONE PROTEINS

the specification of which:

/___/ is attached hereto.

/___/ was filed on ____________ as application Serial No. ____________ and if this box contains an X /___/, was amended on ___________________________.

/ X / was filed as Patent Cooperation Treaty international application No. PCT/US00/21313 on August 3, 2000, if this box contains an X /_/, was amended on under Patent Cooperation Treaty Article 19 on ______ 2001, and if this box contains an X /_/, was amended on _______________.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge my duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim the benefit under Title 35, United States Code, §119 or §365(a)-(b) of any foreign application(s) for patent or inventor's certificate indicated below and of any Patent Cooperation Treaty international applications(s) designating at least one country other than the United States indicated below and have also identified below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application for said subject matter the priority of which is claimed:

Country	Number	Filing Date	Priority Claimed
________	________	________	// Yes // No
________	________	________	// Yes // No

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)
60/146,908	August 3, 1999	Expired
60/160,924	October 22, 1999	Expired

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of Title 35, United States Code §112, I acknowledge my duty to disclose material information as defined in Title 37 Code of Federal Regulations, §1.56(a) which occurred between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)

I hereby appoint the following:

13

Lucy J. Billings	Reg. No. 36,749
Michael C. Cerrone	Reg. No. 39,132
Diana Hamlet-Cox	Reg. No. 33,302
Richard C. Ekstrom	Reg. No. 37,027
Barrie D. Greene	Reg. No. 46,740
Matthew R. Kaser	Reg. No. 44,817
Lynn E. Murry	Reg. No. 42,918
Shirley A. Recipon	Reg. No. 47,016
Susan K. Sather	Reg. No. 44,316
Michelle M. Stempien	Reg. No. 41,327
David G. Streeter	Reg. No. 43,168
Stephen Todd	Reg. No. 47,139
P. Ben Wang	Reg. No. 41,420

respectively and individually, as my patent attorneys and/or agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith. Please address all communications to:

Docket No.: PF-0728 USN

LEGAL DEPARTMENT
INCYTE GENOMICS, INC.
3160 PORTER DRIVE, PALO ALTO, CA 94304

TEL: 650-855-0555 FAX: 650-849-8886 or 650-845-4166

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

1-00

First Joint Inventor:

Full name:	Henry Yue
Signature:	Henry Yue
Date:	September 24, 2001
Citizenship:	United States
Residence:	Sunnyvale, California CA
P.O. Address:	826 Lois Avenue Sunnyvale, California 94087

2-00

Second Joint Inventor:

Full name:	Olga Bandman
Signature:	Olga Bandman
Date:	12 September, 2001
Citizenship:	United States
Residence:	Mountain View, California CA
P.O. Address:	366 Anna Avenue Mountain View, California 94043

3-00

Third Joint Inventor:

Full name: Y. Tom Tang

Signature:

Date: Sept. 10, 2001

Citizenship: United States

Residence: San Jose, California CA

P.O. Address: 4230 Ranwick Court
San Jose, California 95118

4-00

Fourth Joint Inventor:

Full name: Mariah R. Baughn

Signature:

Date: September 5, 2001

Citizenship: United States

Residence: San Leandro, California CA

P.O. Address: 14244 Santiago Road
San Leandro, California 94577

5-00

Fifth Joint Inventor:

Full name: Yalda Azimzai

Signature:

Date: September 13, 2001

Citizenship: United States

Residence: Castro Valley, California

P.O. Address: 5518 Boulder Canyon Drive
Castro Valley, California 94552

Sixth Joint Inventor:

Full name:	Dyung Aina M. Lu
Signature:	Aina Lu
Date:	Sept 7th, 2001
Citizenship:	United States
Residence:	San Jose, California CA
P.O. Address:	233 Coy Drive San Jose, California 95123